Regulation of Pol II transcription and mRNA capping

Kyle Andrew Nilson
University of Iowa

Copyright © 2016 Kyle Andrew Nilson

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/5584

Recommended Citation
https://doi.org/10.17077/etd.3xarspzz

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Cell Biology Commons
REGULATION OF POL II TRANSCRIPTION AND MRNA CAPPING

by

Kyle Andrew Nilson

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Molecular and Cellular Biology in the Graduate College of The University of Iowa

May 2016

Thesis Supervisor: Professor David H. Price
This is to certify that the Ph.D. thesis of

Kyle Andrew Nilson

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Molecular and Cellular Biology at the May 2016 graduation.

Thesis Committee:

David H. Price, Thesis Supervisor

Catherine A. Musselman

Chioma M. Okeoma

Miles A. Pufall

Daniel L. Weeks
ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Dr. David Price for his guidance and support throughout my graduate career. His advice was always blunt and timely, but his patience was near limitless. He required only precision and enthusiasm of his students and as a mentor, he fostered these qualities with great effectiveness.

I would also like to thank past and current members of the Price Lab for creating and maintaining a unique research environment. Dr. Jiannan Guo properly introduced me to biochemistry and taught me how to perform in vitro transcriptions. John Brogie regularly challenged my sanity and provided hours of discussion.

I would like to thank my neighbor and friend Dr. Ruth Spinks for her encouragement through difficult and challenging times. Without her support, this thesis would not exist.

I would like to thank the University of Iowa Graduate College for awarding me the Presidential Graduate Research Fellowship, which gave me the freedom to pursue ambitious and expensive experiments even when funding was questionable.

Finally, I would like to thank my parents Donald and Kathleen Nilson. They are the only mentors I know with more patience than Dr. Price.
ABSTRACT

In humans, RNA polymerase II is the sole source of messenger RNAs that are ultimately translated into proteins and its transcriptional activity is highly regulated. Mechanisms have evolved to control which, when, and to what degree genes are transcribed. Because most cells have the same genome, control of transcription is essential in maintaining cellular identity. Misregulation of Pol II transcription is a hallmark of both cancer and retroviral infection. This research investigates the regulation of Pol II transcription and related co-transcriptional mRNA capping.

Chromatin immunoprecipitation experiments were used to characterize the composition of nucleosomes and Pol II, DSIF and NELF occupancies at bidirectional promoters and enhancers. In collaboration with Alberto Bosque and Vicente Planelles, sequencing experiments were performed in a primary T cell model of HIV latency and a role for sequence-specific recruitment of STAT5 was established in HIV reactivation. In contrast, analysis of Myc binding in vitro and in cells demonstrated that transcription machinery played a major role in recruiting Myc to genomic sites. A precise method was also developed to detect polymerase-associated nascent transcripts in nuclei.

The roles of Cdk7, a subunit of TFIIH that phosphorylates Pol II during initiation, were characterized by treatment of nuclear extracts and cells with THZ1, a recently developed covalent inhibitor with anti-cancer properties. Inhibition of Cdk7 was demonstrated to cause defects in Pol II phosphorylation, co-transcriptional capping, promoter proximal pausing, and productive elongation. Capping of nascent RNAs was found to be spatially and temporally regulated in part by a previously undescribed THZ1-sensitive factor present in nuclear extract. THZ1 impacted pausing through a capping-independent block of DSIF and NELF loading. The P-TEFb-dependent transition into productive elongation was also inhibited by THZ1, likely due to misloading of DSIF.

In vitro and sequencing methods were used to describe an extremely rapid and global transcriptional response to hydrogen peroxide. During periods of oxidative stress, termination was likely inhibited and Pol II accumulated at promoters and enhancers after as few as two minutes, and clearance of these polymerases required P-TEFb. In the presence of flavopiridol, a potent P-TEFb inhibitor, non-productive elongation was observed and a potential role for P-TEFb in termination was proposed.
PUBLIC ABSTRACT

Humans are composed of millions of cells and most of these have the same DNA, or blueprint. Different cells transcribe different RNAs, or messages, to create different proteins, which perform the work of the cell. Control of this transcription is what makes these different cells within different tissues and organs function, and breakdown of this control leads to developmental defects and cancer. HIV also hijacks this control and this makes AIDS difficult to cure. During my graduate career, I developed methods to study this control genome-wide and characterized how a potentially new HIV drug works. By using small-picture biochemistry assays to study THZ1, a new cancer drug, I also identified new ways in which transcription of these messages are controlled. Finally, I discovered rapid transcriptional response to hydrogen peroxide. Overall, this research demonstrates the importance of controlling transcription both generally and in response to cellular stress.
TABLE OF CONTENTS

LIST OF FIGURES ......................................................................................................................... vii
LIST OF ABBREVIATIONS ............................................................................................................ ix

CHAPTER 1: INTRODUCTION ....................................................................................................... 1
  Properties of RNA Polymerase II ................................................................................... 2
  Transcription Elongation Control ................................................................................. 6
  Co-Transcriptional mRNA Capping ............................................................................. 10
  Transcription in Three Dimensions: Chromatin and Enhancers ............................ 12
  Focus of the Thesis ......................................................................................................... 16

CHAPTER 2: GENOME-WIDE STUDIES OF ELONGATION CONTROL ..................................... 17
  Introduction................................................................................................................... 17
  Materials and Methods ................................................................................................. 19
    ChIP-Seq ............................................................................................................. 19
    RNA-Seq .............................................................................................................. 25
    Myc occupancy analyses .................................................................................. 26
    Nuclear walk-on ................................................................................................ 26
    In vitro transcription ....................................................................................... 27
    Transcript separation techniques .................................................................. 28
    CIP and HCE add-backs ..................................................................................... 29
  Results .............................................................................................................................. 30
    Optimization of chromatin immunoprecipitation ...................................... 30
    Elongation control at bidirectional promoters and enhancers .......... 36
    Transcription in a primary T cell model of HIV latency ............................. 44
    Targeted reactivation of HIV by STAT5 after HODHBt treatment ......... 48
    Possible regulation of HEXIM1 by transcriptional interference ............... 52
    Sequence specificity poorly defines the genomic occupancy of Myc ...... 58
    Methods to detect and isolate Pol II nascent transcripts ....................... 61
  Discussion ......................................................................................................................... 70

CHAPTER 3: CDK7 LINKS CAPPING AND ELONGATION CONTROL ..................................... 74
  Introduction................................................................................................................... 74
  Materials and Methods ................................................................................................. 76
    In vitro transcription ....................................................................................... 76
    Kinase assay ....................................................................................................... 77
    Factor add-backs ............................................................................................... 77
    Transcript cap status determination ............................................................. 78
    Nuclear run-on .................................................................................................. 79
    Nuclear walk-on ................................................................................................ 80
  Results .............................................................................................................................. 80
    Effects of THZ1 on transcription in vitro ...................................................... 80
    THZ1 inhibits early, efficient capping ........................................................... 82
    THZ1 effects on pausing are due to defective recruitment of DSIF and
    NELF, not impaired capping ............................................................................ 87
    Effects of THZ1 seen in vitro also occur in cells ................................. 92
Discussion........................................................................................................................................96

CHAPTER 4: TRANSCRIPTIONAL RESPONSES TO OXIDATIVE STRESS........................................100

Introduction.......................................................................................................................................100
Materials and Methods .................................................................................................................102
  In vitro transcription ..................................................................................................................102
  Nuclear walk-on .....................................................................................................................102
  ChIP-Seq ....................................................................................................................................103

Results...............................................................................................................................................105
  Hydrogen peroxide inhibits capping, productive elongation, and
  initiation in vitro ......................................................................................................................105
  H₂O₂ induces a global transcriptional response in cells .......................................................107
  P-TEFb contributes to a rapid response to H₂O₂ .................................................................111
  Nascent transcript cap status after H₂O₂ treatment .............................................................112
  H₂O₂ alters Pol II occupancy genome-wide ........................................................................116
  The role of P-TEFb at promoters and enhancers ..............................................................120

Discussion.......................................................................................................................................125

CHAPTER 5: SUMMARY AND FUTURE DIRECTIONS .................................................................130

REFERENCES.................................................................................................................................135
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regulation of Pol II transcription and mRNA capping</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Recruitment of P-TEFb by Tat from the 7SK snRNP</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>Chromatin around promoters and enhancers</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>An overview of ChIP-Seq</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>Tris, not glycine, quenches formaldehyde</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>ChIP-qPCR signal-to-noise at Myc promoter using various crosslinkers</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>Paired-end reads improve ChIP-Seq resolution around promoters</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>Metagene analysis of elongation machinery and histone modifications</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td>Heatmap analysis of elongation machinery and histone modifications</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>Elongation machinery and chromatin marks around promoters</td>
<td>41</td>
</tr>
<tr>
<td>11</td>
<td>Elongation machinery and chromatin marks around enhancers</td>
<td>43</td>
</tr>
<tr>
<td>12</td>
<td>Pol II occupancy at genes differentially expressed in T cells ± αCD3/αCD28</td>
<td>47</td>
</tr>
<tr>
<td>13</td>
<td>ChIP-Seq occupancy of Pol II over the HIV genome ± αCD3/αCD28</td>
<td>48</td>
</tr>
<tr>
<td>14</td>
<td>ChIP-Seq occupancies of Pol II and STAT5A over the HIV genome</td>
<td>49</td>
</tr>
<tr>
<td>15</td>
<td>Differential expression in T cells after HODHBt or αCD3/αCD28</td>
<td>51</td>
</tr>
<tr>
<td>16</td>
<td>Differential expression of genes significantly changed by HODHBt and/or αCD3/αCD28</td>
<td>51</td>
</tr>
<tr>
<td>17</td>
<td>RNA-Seq signal over genes significantly changed only by HODHBt</td>
<td>53</td>
</tr>
<tr>
<td>18</td>
<td>RNA-Seq signal over genes significantly changed only by αCD3/αCD28</td>
<td>54</td>
</tr>
<tr>
<td>19</td>
<td>RNA-Seq signal over genes significantly changed by both treatments</td>
<td>55</td>
</tr>
<tr>
<td>20</td>
<td>Possible regulation of HEXIM1 gene expression by antisense transcription</td>
<td>57</td>
</tr>
<tr>
<td>21</td>
<td>The spectrum of Myc-Max in vitro binding affinities to all possible 8-mers</td>
<td>59</td>
</tr>
<tr>
<td>22</td>
<td>Sites of Myc occupancy in cells are slightly enriched for quality E-boxes</td>
<td>60</td>
</tr>
<tr>
<td>23</td>
<td>Myc ChIP-Seq occupancy levels weakly correlate with E-box quality</td>
<td>62</td>
</tr>
<tr>
<td>24</td>
<td>Nuclear run-ons with α-³²P-CTP only, biotin-11-NTPs, or BrUTP</td>
<td>64</td>
</tr>
<tr>
<td>25</td>
<td>Incorporation of biotin-11-NTPs by in vitro elongation complexes</td>
<td>66</td>
</tr>
<tr>
<td>26</td>
<td>Responsiveness of transcripts to phosphatase or capping enzyme</td>
<td>69</td>
</tr>
<tr>
<td>27</td>
<td>Timing of effects of THZ1 on in vitro transcription</td>
<td>82</td>
</tr>
<tr>
<td>28</td>
<td>Effects of THZ1 and flavopiridol on P-TEFb kinase activity</td>
<td>83</td>
</tr>
<tr>
<td>29</td>
<td>THZ1-treated elongation complexes respond differently to HCE</td>
<td>84</td>
</tr>
<tr>
<td>30</td>
<td>Activity of HCE added to high salt washed elongation complexes ± THZ1</td>
<td>85</td>
</tr>
<tr>
<td>31</td>
<td>Activity of HCE added to low salt washed elongation complexes ± THZ1</td>
<td>86</td>
</tr>
</tbody>
</table>
Figure 32. Activity of HCE added to preinitiation complexes ± THZ1 ........................................ 87
Figure 33. THZ1-induced pause defect is independent of mRNA capping ................................. 88
Figure 34. Effects of H₂O₂ on HCE added to high salt washed elongation complexes ......... 89
Figure 35. Pausing factor add-backs to high or low salt washed complexes ± THZ1....... 90
Figure 36. Profiles of pausing factor add-backs to low salt washed complexes ± THZ1 ... 91
Figure 37. Cap status of transcripts in cells treated with DMSO, flavopiridol, or THZ1 .. 93
Figure 38. Profiles of total and capped Pol II transcripts in cells treated with DMSO, flavopiridol, or THZ1 .................................................................................................................. 94
Figure 39. THZ1 inhibits proper mRNA capping and Pol II elongation in cells .......... 95
Figure 40. The role of Cdk7 in CTD phosphorylation, capping, pausing, and productive elongation .................................................................................................................................... 97
Figure 41. Effects of H₂O₂ on transcription in vitro ............................................................. 106
Figure 42. Effects of H₂O₂ on global transcription detected by nuclear walk-on.......... 108
Figure 43. Effects of H₂O₂ in the absence and presence of flavopiridol ......................... 110
Figure 44. Cap status of nascent transcripts after H₂O₂ treatment ............................... 113
Figure 45. Profiles of total and capped nascent transcripts after H₂O₂ treatment .... 114
Figure 46. Metagene analysis of Pol II occupancy in H₂O₂-treated cells ......................... 117
Figure 47. Heatmap analysis of Pol II occupancy in H₂O₂-treated cells ......................... 119
Figure 48. Pol II occupancy over HSPA5 in H₂O₂-treated cells ± flavopiridol .......... 121
Figure 49. Pol II occupancy over FOS in H₂O₂-treated cells ± flavopiridol ................. 122
Figure 50. ChIP-Seq occupancy around JUNB enhancers after H₂O₂ ± flavopiridol ....... 123
Figure 51. ChIP-Seq occupancy around DDIT4 enhancers after H₂O₂ ± flavopiridol ...... 124
Figure 52. Resolution of bidirectional pausing with P-TEFb ± H₂O₂ ............................... 127
LIST OF ABBREVIATIONS

ADP ........................................................................................................... adenosine diphosphate
AIDS ............................................................................. acquired immune deficiency syndrome
ATP .......................................................................................................... adenosine triphosphate
BrUTP ....................................................................................... 5-bromouridine 5′-triphosphate
BSA ............................................................................................................ bovine serum albumin
CBC  ............................................................................................... cap-binding protein complex
CDK ........................................................................................................ cyclin-dependent kinase
ChIP ......................................................................................... chromatin immunoprecipitation
ChIP-Seq ........................................................................................ ChIP followed by sequencing
CIP ...................................................................................... calf intestinal alkaline phosphatase
CMV ..................................................................................................................... cytomegalovirus
CTD ...................................................................................................... carboxy-terminal domain
CTP ............................................................................................................. cytidine triphosphate
DMEM ...................................................................................... Dulbecco’s modified Eagle media
DMSO ............................................................................................................... dimethyl sulfoxide
DNA ............................................................................................................ deoxyribonucleic acid
DNase .............................................................................................. deoxyribonuclease
DRB .............................................................. 5,6-dichloro-1-β-D-ribofuranosyl-benzimidazole
ds ......................................................................................................................... double-stranded
DSIF ............................................................................................ DRB sensitivity inducing factor
DTT ............................................................................................................................ dithiothreitol
EC .................................................................................................................... elongation complex
EDTA ........................................................................................ ethylenediaminetetraacetic acid
EGS ........................................................................ ethylene glycol bis(succinimidyl succinate)
FBS .................................................................................................................... fetal bovine serum
GRO-Seq .......................................................... global nuclear run-on followed by sequencing
GTF .................................................................................................. general transcription factor
GTP ..........................................................................................................guanosine triphosphate
H3K27ac ...................................................................................... histone 3 lysine 27 acetylation
H3K4me1 .................................................................................... histone 3 lysine 4 monomethylation
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K4me3</td>
<td>histone 3 lysine 4 trimethylation</td>
</tr>
<tr>
<td>HCE</td>
<td>human capping enzyme</td>
</tr>
<tr>
<td>HCM</td>
<td>human cap methyltransferase</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>HEXIM</td>
<td>hexamethylene bisacetamide-inducible protein</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HNE</td>
<td>HeLa nuclear extract</td>
</tr>
<tr>
<td>HODHBt</td>
<td>3-hydroxy-1,2,3-benzotriazin-4(3H)-one</td>
</tr>
<tr>
<td>HSW</td>
<td>high salt wash</td>
</tr>
<tr>
<td>INR</td>
<td>initiator</td>
</tr>
<tr>
<td>IncRNA</td>
<td>long non-coding RNA</td>
</tr>
<tr>
<td>LSW</td>
<td>low salt wash</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>m7G</td>
<td>7-methylguanosine</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro RNA</td>
</tr>
<tr>
<td>MNase</td>
<td>micrococcal nuclease</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NELF</td>
<td>negative elongation factor</td>
</tr>
<tr>
<td>NER</td>
<td>nucleotide excision repair</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleoside triphosphate</td>
</tr>
<tr>
<td>PAF1</td>
<td>Pol II-associated factor 1</td>
</tr>
<tr>
<td>PAF1C</td>
<td>PAF1 complex</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PAR</td>
<td>poly(ADP-ribose)</td>
</tr>
<tr>
<td>PBM</td>
<td>protein-binding microarray</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PIC</td>
<td>preinitiation complex</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>Pol I</td>
<td>RNA polymerase I</td>
</tr>
</tbody>
</table>
Pol II ........................................................................................................................RNA polymerase II
Pol III ......................................................................................................................RNA polymerase III
PRO-Seq ..............................................................................................................precision nuclear run-on followed by sequencing
P-TEFb ..................................................................................................................positive transcription elongation factor b
qPCR ......................................................................................................................quantitative polymerase chain reaction
RIPA .....................................................................................................................radioimmunoprecipitation assay
RNA .....................................................................................................................ribonucleic acid
RNAi .....................................................................................................................RNA interference
RNase ....................................................................................................................ribonuclease
ROC .....................................................................................................................receiver operating characteristic
RT .........................................................................................................................room temperature
RT-qPCR ............................................................................................................reverse transcription followed by qPCR
SAM ......................................................................................................................S-adenosyl methionine
SDS ......................................................................................................................sodium dodecyl sulfate
SEC ......................................................................................................................super elongation complex
shRNA ..................................................................................................................short hairpin RNA
SMEM ..................................................................................................................suspension minimum essential media
snoRNA ...............................................................................................................small nucleolar RNA
snRNA ..................................................................................................................small nuclear RNA
snRNP ..................................................................................................................small nuclear ribonucleoprotein
ss ..............................................................................................................................single-stranded
STAT ....................................................................................................................signal transducer and activator of transcription
SUMO ...................................................................................................................small ubiquitin-like modifier
TAF .........................................................................................................................TBP-associated factor
Tat .........................................................................................................................trans-activator of HIV transcription
TBP .......................................................................................................................TATA-binding protein
tRNA ....................................................................................................................transfer RNA
TSS .........................................................................................................................transcription start site
UTP .......................................................................................................................uridine triphosphate
CHAPTER 1: INTRODUCTION

Most cells in a human have the same genome and this blueprint is transcribed into RNA by RNA polymerases I, II, and III. While Pol I and Pol III are mainly responsible for the synthesis of structural and regulatory nuclear RNAs, Pol II is the sole source of messenger RNAs that are ultimately translated into proteins. The selective expression of genes by Pol II is essential for differentiation and maintenance of cellular identity, without which cooperativity between specialized cells, tissues, and organs would not be possible. Mechanisms have evolved to control which, when, and to what degree genes are transcribed and regulatory targets include the polymerase itself and its associated nascent transcripts, and the chromatin surrounding these genes. Misregulation of Pol II transcription results in developmental defects including cancer and is a hallmark of retroviral infection. Despite decades of progress, additional layers of regulation continue to be discovered in this fundamental process and research of these mechanisms remains a top priority.

From start to finish, transcription is highly regulated and a number of coordinated events are required to generate a fully processed transcript. Pol II is generally observed in one of four transcriptional phases: preinitiation complex (PIC) formation, promoter-proximal pausing, productive elongation, and termination (Figure 1). PIC formation requires the coordinated assembly of Pol II subunits and the general transcription factors TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Sainsbury et al., 2015). Select residues within the Pol II C-terminal domain (CTD) are also phosphorylated during initiation and this phosphorylation changes throughout transcription (Heidemann et al., 2013). Just downstream of most transcription start sites, pausing is induced by the DRB sensitivity inducing factor DSIF in conjunction with the negative elongation factor NELF (Yamaguchi et al., 2013). A significant portion of these polymerases also have Gdown1, which prevents premature termination (Cheng et al., 2012). Release from pausing and the transition into productive elongation is triggered by positive transcription elongation factor b (P-TEFb), whose recruitment and kinase activity are tightly regulated (Guo and Price, 2013). Guanylylation of the nascent transcript by human capping enzyme (HCE), and subsequent methylation by human cap methyltransferase (HCM) occurs early during elongation and in addition to
Figure 1. Regulation of Pol II transcription and mRNA capping
The roles of factors throughout transcription by Pol II (green) are illustrated. Relative C-terminal domain phosphorylation is depicted by color intensity (red > pink > grey).

protecting the transcript, helps to recruit the processing machinery required for successful mRNA production (Bentley, 2014). At the 3’ end of genes, polymerases slow and accumulate downstream of polyadenylation sites, where termination factors ultimately recycle Pol II for future rounds of transcription (Guo and Price, 2013). By controlling every step of transcription, cells are able to finely tune their expression programs and, when needed, selectively and rapidly activate any gene.

Properties of RNA Polymerase II

RNA polymerase II is composed of twelve subunits, Rpb1-12 numbered largest to smallest, which assemble to form a claw which wraps around the DNA. The hinge is formed first by Rpb3, 10, 11, and 12 (Wild and Cramer, 2012). Then, Rpb2 docks with this hinge and, with Rpb9, forms one of the two pincers (Wild and Cramer, 2012). The claw
is completed by association of the largest subunit, Rpb1, along with Rpb5, 6, and 8 (Wild and Cramer, 2012). Rpb1 also has a flexible CTD composed of 52 repeats of the heptad sequence YSPTSPS (Hahn, 2004); select residues within the CTD are dynamically phosphorylated throughout transcription (Heidemann et al., 2013). Opposite of the cleft formed by the two pincers is a stalk composed of Rpb4 and 7, which reversibly associates with the Rpb1 subassembly (Wild and Cramer, 2012). The catalytic core of Pol II is buried at the junction of the Rpb1, 2, and 3 subassemblies (Hahn, 2004). As the polymerase elongates, DNA is guided through the Rpb1/2 cleft into the active center where the strands are separated, secured by the Rpb1 clamp, and routed around the wall formed by Rpb3 (Hahn, 2004). Nucleoside triphosphates (NTPs) enter through a pore in Rpb1 and are coordinated with the template strand to synthesize RNA (Hahn, 2004). The nascent transcript is then routed through an exit pore and guided away from Pol II by the Rpb4/7 stalk (Hahn, 2004). Pol II is closely related not only to Pol I and III, but also to archaeal RNA polymerase. Rpb5, 6, 8, 10, and 12 are all shared between human polymerases but the subunits that form the pincers, the hinge, and the stalk are tailored between them to interact with each’s specialized accessory factors (Vannini and Cramer, 2012).

Initiation of transcription requires TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, and these general transcription factors (GTFs) make contacts throughout the core Pol II structure. Portions of TFIIB extend into the cleft and wall to read and position DNA over the active site (Sainsbury et al., 2015). TFIIB also connects Pol II to TFIID, which is composed of TATA-binding protein (TBP) and several TBP-associated factors (TAFs). TFIIE binds the Rpb1 clamp and is thought to bind to single-stranded DNA (Sainsbury et al., 2015). TFIIF spans across the cleft and interacts with Rpb9, the clamp, and DNA (Sainsbury et al., 2015). TFIIH, a large ten-subunit factor, is recruited by TFIIE and interacts directly with DNA downstream of the cleft and with the Pol II Rpb1 CTD (Sainsbury et al., 2015). Overall, these GTFs modulate and constrain the multiple hinges and moving components in Pol II early in transcription and are considered essential components of the basal transcription machinery. Pol I, II, and III have separate, somewhat conserved factors which emulate TFIIB, TFIIE, and TFIIF (Vannini and Cramer, 2012). All three polymerases use TBP, but only Pol II associates with a greater
TFIID complex (Vannini and Cramer, 2012). Additionally, TFIIH is unique to Pol II and is not required for transcription by Pol I or III (Vannini and Cramer, 2012). Other than Rpb9, TFIIF, and TFIIH, all eukaryotic subunits and GTFs are conserved from archaea (Werner and Grohmann, 2011). Five fundamental Pol II subunits—Rpb1, 2, 3, 11, and 6—also share homology with bacterial RNA polymerase (Decker and Hinton, 2013).

Multisubunit RNA polymerases all share the same mechanism of action, but as their complexity increased over time, they have become more dependent on transcription factors to control initiation, elongation, and termination (Werner and Grohmann, 2011).

Initiation does not occur randomly across the genome; rather, Pol II transcription originates at promoters, regions of DNA which contain sequence elements that interact with the basal transcription machinery (Smale and Kadonaga, 2003). The first core promoter element discovered was the TATA box, which is recognized by TBP. TATA boxes are generally positioned 30 base pairs (bp) upstream from human transcription start sites (TSSs) and TATATAAG is the optimized TBP recognition sequence (Wong and Bateman, 1994). Some TATA boxes also have a TFIIIB recognition element immediately upstream which is bound by a helix-turn-helix motif absent in the yeast protein (Lagrange et al., 1998). Several promoters were found to have a pair of pyrimidines arranged around an adenosine at the +1 position (Corden et al., 1980). This element, later termed initiator (INR), is bound by TAFs within TFIID and acts synergistically with TATA boxes located 25-30 bp upstream (O’Shea-Greenfield and Smale, 1992). The downstream promoter element, conserved from Drosophila melanogaster to humans, is found 28-32 bp downstream of the TSS and, with INR, can functionally substitute for TATA at some promoters (Kutach and Kadonaga, 2000).

The first step in preinitiation complex (PIC) formation is the sequence-specific binding of TBP to the TATA box (if one is present). TBP and TAFs will induce a bend in the DNA and this structure is then stabilized by TFIIA and TFIIIB (Sainsbury et al., 2015). This upstream promoter complex is then joined by Pol II and TFIIF and is subsequently completed by recruitment of TFIIE and TFIIH. In the presence of ATP, the helicase activity of TFIIH will separate the strands of DNA, generating a transcription bubble where nucleoside triphosphates (NTPs) can be rapidly compared against the template...
strand (Vannini and Cramer, 2012). At this point, RNA synthesis occurs and subsequent dissociation of GTFs will lead to promoter clearance, thus marking the end of initiation and the beginning of elongation (Luse, 2013).

Despite the important role of promoter sequences in targeting TFIIB and TFIID during initiation, mammalian promoters frequently lack identifiable core promoter elements. Although TBP is required for initiation, only 10-20% of annotated human promoters contain a functional TATA box (Sandelin et al., 2007). Additionally, only 30% of detected transcription initiation sites in murine macrophages had INR elements and less than 8% had a TATA box upstream despite these promoters having relatively high levels of sequence conservation (Scruggs et al., 2015). TBP can interact with a wide variety of A/T-rich sequences (Wong and Bateman, 1994) and this is thought to enable TBP function within the three eukaryotic polymerases at their respective promoters (Vannini and Cramer, 2012). Because of the loose sequence specificities of core promoter elements, it has been challenging to correlate them in a statistically significant manner with sites of PIC formation (Siebert and Soding, 2014; Venters and Pugh, 2013, 2014). The core promoter elements are, however, found at similar frequencies across all sites of human Pol II initiation, including at enhancers and divergent promoters, suggesting their continued importance in influencing TSS selection (Core et al., 2014; Pugh and Venters, 2016).

During initiation, the TFIIH kinase module composed of Cdk7, Cyclin H, and Mat1 phosphorylates the CTD of the large Rpb1 subunit of Pol II (Glover-Cutter et al., 2009) and this is the first of several CTD phosphorylation and dephosphorylation events during the transcription cycle (Heidemann et al., 2013). Although Cdk7 is thought to primarily phosphorylate Ser5 and Ser7 residues within the 52 heptapeptide CTD, recent evidence suggests it could also phosphorylate Ser2 to a lesser degree within 30 seconds of initiation (Nilson et al., 2015). After the transition into productive elongation, Ser2 becomes the predominantly phosphorylated CTD residue and Cdk9 (Marshall et al., 1996) and Cdk12 (Bartkowiak et al., 2010; Bowman and Kelly, 2014) have been implicated in this phosphorylation. As Pol II travels towards the 3’ ends of genes, a number of phosphatases act to remove phosphates from Ser5 and Ser7 (Bataille et al., 2012; Egloff et al., 2012; Hsu et al., 2014; Mosley et al., 2009; Zhang et al., 2012a; Zhang et
These dynamic changes to the CTD are thought to orchestrate the recruitment of initiation, elongation, and termination factors at appropriate times during transcription (Heidemann et al., 2013).

**Transcription Elongation Control**

In the absence of other factors, elongating Pol II is slow—under 300 bp/min (Izban and Luse, 1992)—and prone to transient pausing and arrest (Guo and Price, 2013). Transient pauses, which vary in their duration, are resolved without additional factors and can be induced by certain DNA sequences, DNA damage, NTP misincorporation into the nascent transcript, or structural elements present in the nascent transcript (Martinez-Rucobo and Cramer, 2013). The combined activities of DSIF and NELF increase the dwell time at these pause sites about 3-fold (Renner et al., 2001; Yamaguchi et al., 2013). In contrast, TFIIF stimulates elongation about 20-fold at these transient pause sequences (Price et al., 1989). In certain conditions, Pol II will backtrack, which causes the 3’ end of the nascent transcript to disengage from the active site of RNA synthesis (Martinez-Rucobo and Cramer, 2013). These polymerases will remain arrested until TFIIS induces a conformational change within Pol II that cleaves the excess length of nascent transcript, thus re-enabling forward motion (Cheung and Cramer, 2011). DSIF and NELF further encourage polymerase arrest by inhibiting transcript cleavage factor TFIIS (Palangat et al., 2005), which may be generally required to restart elongation (Adelman et al., 2005). Notably, TFIIS-dependent cycling of Pol II between backtracking, arrest, and cleavage is widespread downstream of promoters in *Drosophila* (Nechaev et al., 2010).

The selective association of factors results in two distinctly different elongation modes of Pol II. By using an in vitro system where HeLa nuclear extract and a CMV-promoter-driven template are first preincubated to enable PIC formation, and then incubated with NTPs to allow for initiation, it was shown that Pol II normally enters an abortive mode of elongation characterized by the generation of short transcripts, pausing, and termination (Marshall and Price, 1992). Promoter-proximal pausing was later shown to require DSIF (Wada et al., 1998) and NELF (Yamaguchi et al., 1999), which cooperatively reduce the Pol II elongation rate (Renner et al., 2001). Paused polymerases which contain the substoichiometric Pol II subunit Gdown1 are also
resistant to both elongation stimulation by TFIIF and termination by TTF2 (Cheng et al., 2012; Guo et al., 2014b; Mullen Davis et al., 2014). In a factor-dependent manner, these abortive elongation complexes could transition into a mode of “productive elongation” and generate full-length transcripts (Marshall and Price, 1992); this factor was later purified and named positive transcription elongation factor b (P-TEFb) (Marshall and Price, 1995). During this transition into productive elongation, the Cdk9 kinase subunit of P-TEFb phosphorylates DSIF, which remains associated with Pol II and subsequently acts as a positive elongation factor (Bernecky et al., 2016; Yamada et al., 2006). NELF is also phosphorylated and dissociates from elongation complexes (Fujinaga et al., 2004). This method of regulating Pol II elongation explained previous findings where repression of Myc expression after differentiation (Bentley and Groudine, 1986) and activation of both human immunodeficiency virus (HIV) by Tat (Kao et al., 1987) and HSP70 in Drosophila (O'Brien and Lis, 1991; Rougvie and Lis, 1988, 1990) involved polymerases that had already initiated and were transcriptionally engaged, but paused. Promoter-proximal pausing of Pol II by DSIF and NELF has since been shown to occur on essentially all expressed genes in metazoans (Rahl et al., 2010).

Because promoter-proximal pausing of Pol II is widespread, the activity of P-TEFb must be tightly controlled and this is accomplished by its reversible association with the 7SK small nuclear ribonucleoprotein (snRNP) complex (Peterlin et al., 2012). The protein composition of 7SK snRNP is dynamic. Although LARP7 (Krueger et al., 2008) and MePCE (Xue et al., 2010) are constitutively associated with 7SK and aid in its assembly (Muniz et al., 2013), P-TEFb can only enter snRNPs that contain HEXIM1 (Barboric et al., 2005; Michels et al., 2004) or HEXIM2 (Byers et al., 2005), which inhibit P-TEFb. Notably, only P-TEFb whose Cdk9 subunit is phosphorylated (and thus kinase active) can associate with 7SK (Chen et al., 2004; Li et al., 2005). When P-TEFb is eventually recruited and released, a conformational change in the snRNP ejects HEXIM proteins (Krueger et al., 2010) and this form of 7SK is stabilized by hnRNPs (Krueger et al., 2008). A functional equilibrium is maintained between free and sequestered P-TEFb to maintain transcriptional levels within cells. When transcription is globally disrupted, P-TEFb is rapidly released (Bartholomeeusen et al., 2013; Bartholomeeusen et al., 2012; He et al., 2006; Krueger et al., 2008). The balance between small and large forms of the
7SK snRNP is restored by subsequent induction of *HEXIM1* expression (Liu et al., 2014) and reduction of P-TEFb protein levels (Krueger et al., 2008).

To enable productive elongation, P-TEFb is recruited from the 7SK snRNP by a number of cellular activators including Brd4 (Krueger et al., 2010; Yang et al., 2005), c-Myc (Eberhardy and Farnham, 2002; Kanazawa et al., 2003; Rahl et al., 2010), NFκB (Barboric et al., 2001), and others (Peterlin and Price, 2006). The most potent recruiter currently known, however, is the transactivator of HIV transcription protein Tat (Nilson and Price, 2011). Tat directly interacts with P-TEFb (Tahirov et al., 2010), removes it from the 7SK snRNP (Barboric et al., 2007; Krueger et al., 2010; Sedore et al., 2007), and recruits it to the HIV promoter through an interaction with TAR, the transactivation response element formed by the viral nascent transcript (Jones and Peterlin, 1994; Wei et al., 1998) (Figure 2). In the presence of Tat, cellular control of P-TEFb via the 7SK snRNP is no longer effective and this guarantees a supply of P-TEFb for HIV replication. At the most highly expressed genes, P-TEFb is found in association with the super elongation complex (SEC), which is variably composed of a number of elongation factors including ELL, ENL, AFF, and EAF proteins (Luo et al., 2012). Notably, the SEC interacts with Brd4 (Flajollet et al., 2013; Fowler et al., 2014) and is directly recruited by Tat to the HIV genome (Chou et al., 2013; Gu et al., 2014; He et al., 2010; Lu et al., 2014; Schulze-Gahmen et al., 2014; Schulze-Gahmen et al., 2013; Sobhian et al., 2010). Productive elongation is also facilitated by the Pol II-associated factor 1 (PAF1) complex (Crisucci and Arndt, 2011), although this role has recently been contested (Chen et al., 2015; Yu et al., 2015). The PAF1 complex acts synergistically with phosphorylated DSIF to stimulate elongation in vitro and in vivo (Chen et al., 2009) and is thought to link the SEC to productively elongating Pol II (He et al., 2011). Ultimately, productively elongating Pol II will accelerate over the first 15 kb, probably while loading PAF1C, SEC, and other elongation factors, and travel at an average rate of 3 kb/min until mRNA synthesis is complete (Danko et al., 2013).

Early transcription is also governed by Mediator, a large protein complex whose 26 interchangeable subunits enable variable responsiveness to different transcription factors (Allen and Taatjes, 2015). Mediator makes extensive contacts with Pol II during PIC assembly (Bernecky et al., 2011; Plaschka et al., 2015). Cellular activator proteins,
Figure 2. Recruitment of P-TEFb by Tat from the 7SK snRNP
Illustrated is the removal of P-TEFb by HIV Tat from the 7SK snRNP. Tat interacts with TAR, a structured element in the viral nascent transcript. Recruitment of P-TEFb by Tat leads to viral expression. Such as SREBP-1a and VP16, can significantly alter the structure of Mediator without changing its subunit composition (Ebmeier and Taatjes, 2010) and play roles in orienting Pol II and TFIIF at promoters during initiation (Bernecky et al., 2011; Bernecky and Taatjes, 2012). Phosphorylation of the Pol II CTD plays a key role in breaking Mediator-Pol II contacts and forces rearrangement of the Mediator complex between initiation and elongation (Svejstrup et al., 1997). Recent studies in yeast indicate that CTD phosphorylation by Kin28 (the Cdk7 homolog) influences Mediator and is required for promoter escape after initiation (Jeronimo and Robert, 2014; Wong et al., 2014). This Cdk7-dependent factor exchange between initiation and elongation was strongly suggested to involve Mediator in experiments using human nuclear extracts (Nilson et al., 2015). Rearrangement of Mediator is thought to help Mediator accomplish different tasks during initiation and elongation (Poss et al., 2013). For example, PIC-bound Mediator structurally excludes the Cdk8 submodule (Cdk8, Cyclin C, Med12, and Med13) (Bernecky et al., 2011), whose activity is required after initiation.
for recruitment of P-TEFb and Brd4 to serum-responsive immediate-early genes (Donner et al., 2010). Curiously, Cdk8-Mediator does not contain Med26 (Ebmeier and Taatjes, 2010), which interacts with both TFIID during initiation and SEC during elongation (Takahashi et al., 2011). Mediator has many more roles including potential regulation of Gdown1 (Li and Price, 2012; Wu et al., 2012), co-transcriptional capping (Nilson et al., 2015) and splicing (Huang et al., 2012; Huang et al., 2015), and Cohesin-mediated chromatin looping between enhancers and promoters (Dowen et al., 2014; Ji et al., 2016; Kagey et al., 2010; Schaaf et al., 2013).

**Co-Transcriptional mRNA Capping**

As Pol II productively elongates, a number of mRNA processing factors associate with the transcription complex and act to modify the nascent transcript. Most protein-coding transcripts will gain a 7-methylguanosine (m⁷G) cap, have their introns excised and exons ligated during splicing, and undergo 3′ cleavage and polyadenylation to generate mRNAs ready to export to the cytoplasm for translation (Bentley, 2014). Co-transcriptional mRNA processing is exclusive to Pol II (Sisodia et al., 1987) and it was thought early on that the phosphorylated CTD acted as a landing pad for mRNA processing machinery (Greenleaf, 1993). Later studies showed that capping (Cho et al., 1997; McCracken et al., 1997a; Yue et al., 1997), splicing (Hirose et al., 1999), and polyadenylation (Hirose and Manley, 1998; McCracken et al., 1997b) were all stimulated by CTD phosphorylation, albeit to varying degrees. Capping and subsequent cap methylation are essential for successful transcription for two key reasons: 1) uncapped transcripts are rapidly degraded by 5′-to-3′ exoribonucleases, and 2) splicing, polyadenylation, and nuclear export all depend on cap-binding protein complex (CBC) which recognizes m⁷G-capped transcripts (Ghosh and Lima, 2010). CBC interacts directly with P-TEFb and CBC depletion attenuated both HIV transactivation by Tat and alternative splicing (Lenasi et al., 2011). Capping is also required for nascent transcript cleavage (Adamson et al., 2005) and termination is regulated in part by CBC (Porrua and Libri, 2015).

Capping is the first of the linked co-transcriptional processing events required to generate a functional mRNA (Bentley, 2014). Capping occurs in two enzymatic steps. First, RNA triphosphatase removes the terminal phosphate from the 5′ end of the
nascent transcript. Then, RNA guanylyltransferase mediates the 5′–5′ addition of GMP, forming the GpppN cap structure. In yeast, this is accomplished with two enzymes: Cet1 and Ceg1; in humans, these activities are combined into one capping enzyme (HCE) encoded by RNGTT. Guanylylation of nascent transcripts can occur as soon as the RNA emerges from Pol II (Coppola et al., 1983; Moteki and Price, 2002; Rasmussen and Lis, 1993) and a recent structure shows that yeast capping enzymes are positioned immediately adjacent to the RNA exit channel (Martinez-Rucobo et al., 2015). While capping enzymes associate with the phosphorylated Pol II CTD (Chiu et al., 2002; Cho et al., 1997; McCracken et al., 1997a; Yue et al., 1997), the site of this interaction is not conserved between yeast and mammalian capping enzymes (Ghosh et al., 2011). In contrast with fission yeast, where an otherwise lethal Ser5Ala CTD mutation was rescued by fusing mouse capping enzyme to the CTD (Schwer and Shuman, 2011), CTD phosphorylation has only a 4-fold influence on mammalian capping of soluble RNA (Ho and Shuman, 1999) or nascent Pol II transcripts (Mandal et al., 2004; Moteki and Price, 2002; Nilson et al., 2015). HIV Tat has also been shown to stimulate co-transcriptional capping by influencing Pol II CTD phosphorylation (Zhou et al., 2003) and by direct interaction with the mammalian capping enzyme (Chiu et al., 2001; Chiu et al., 2002). Capping enzymes interact with the Spt5 subunit of DSIF in yeast (Doamekpor et al., 2014; Doamekpor et al., 2015; Pei and Shuman, 2002; Schneider et al., 2010) and human DSIF causes a 2 to 5-fold stimulation of co-transcriptional capping (Mandal et al., 2004; Wen and Shatkin, 1999). Additionally, wild type or capping-defective HCE relieved the negative influence of NELF during transcription in vitro suggesting a competition of HCE and pausing factors for Pol II (Mandal et al., 2004). Although yeast do not exhibit promoter proximal pausing, this network of interactions among the capping enzyme, the Pol II CTD, and DSIF could play a role in controlling capping and elongation in mammals.

Cap methylation occurs separately from guanylylation and the final m7G cap influences splicing, polyadenylation, and nuclear export through interactions with CBC, and translation through interaction with eIF4E (Topisirovic et al., 2011). Human cap methyltransferase, encoded by RNMT, acts exclusively on the N7 position of guanylylated RNA and uses S-adenosyl methionine as a methyl donor. Because this
methylation is not easily reversible, cap methylation is thought to “lock in” the cap structure (Dunn and Cowling, 2015). In humans, cap methyltransferase associates with RAM, or RNMT-Activating Mini protein (previously Fam103a1) as a potentially required co-factor (Gonatopoulos-Pournatzis et al., 2011). The N-terminal domain of HCM is required for recruitment to elongation complexes (Aregger and Cowling, 2013) and is phosphorylated by Cdk1-Cyclin B1 during mitosis, which increases methyltransferase activity at the beginning of G1 (Aregger et al., 2016). In budding yeast, cap methyltransferase Abd1 associates with the Pol II CTD and unlike Ceg1, can travel with elongating polymerases to the 3’ end of genes (Lidschreiber et al., 2013; Schroeder et al., 2000). Cap methyltransferase interacts with and potentially recruits Cdk9 in both budding (Lidschreiber et al., 2013) and fission yeast (Guiguen et al., 2007; St Amour et al., 2012). However, recruitment of HCM to promoters is inhibited by DRB, a kinase inhibitor that blocks both CTD phosphorylation and P-TEFb activity (Aregger and Cowling, 2013). c-Myc, a cellular recruiter of P-TEFb, also promotes cap methylation of its target genes (Cowling and Cole, 2007). Because this Myc-induced cap maturation is greater than the Myc-induced increase in transcription (Cole and Cowling, 2009), it is possible that cap methylation might not be a constitutive “housekeeping” event in transcription. Myc also induces expression of S-adenosyl homocysteine hydrolase, which breaks down the inhibitory byproduct of cap methyltransferase activity (Fernandez-Sanchez et al., 2009). Cap methylation levels and NELF occupancy were also recently correlated with nucleosome-induced pausing (Jimeno-Gonzalez et al., 2015). Overall, these studies suggest that simultaneous stimulation of cap maturation and productive elongation may be a powerful means of promoting gene expression and protein production (Cowling and Cole, 2010; Dunn and Cowling, 2015).

**Transcription in Three Dimensions: Chromatin and Enhancers**

Transcription takes place in the context of chromatin, the dynamic 3D structure of DNA tightly wound around nucleosomes within the nucleus. Nucleosomes are generally composed of two copies of each of these canonical histones: H3, H4, H2A, and H2B (Kouzarides, 2007). Although unmodified nucleosomes block Pol II elongation in vitro (Bondarenko et al., 2006), the substitution and modification of individual histones before, during, and after transcription adds a new epigenetic layer of control and
facilitates selective gene expression (Venkatesh and Workman, 2015). Histones are post-translationally modified by a number of transcriptional regulators and chromatin remodelers in response to developmental cues, signaling pathways, cellular stress, and DNA damage (Tessarz and Kouzarides, 2014). Pol II pausing is thought to keep promoter regions open and available for rapid induction of transcription (Gilchrist et al., 2010; Gilchrist et al., 2008). Promoter proximal pausing, however, could be influenced by the positioning of nucleosomes immediately downstream of the transcription start site (Jimeno-Gonzalez et al., 2015). Due to difficulties in studying transcriptional dynamics through chromatin in vitro, only a select few chromatin marks have been “decoded” and their roles in elongation remain somewhat unclear.

One histone modification associated with productive elongation is histone 3 lysine 4 trimethylation (or H3K4me3). In yeast, H3K4 methylation mono-, di-, and trimethylation are carried out by Set1 in association with the COMPASS complex (Krogan et al., 2002; Miller et al., 2001; Roguev et al., 2001) and recruitment of COMPASS to promoters (and subsequent H3K4 methylation) requires the Paf1 complex (Krogan et al., 2003; Mueller et al., 2006; Schneider et al., 2005). While yeast lack promoter-proximal pausing and NELF, Paf1 recruitment and activity were linked with phosphorylation of Spt4, Spt5, and the Rpb1 CTD (Mayekar et al., 2013; Qiu et al., 2012; Qiu et al., 2006). The activity of Ctk1, the yeast homolog of human Cdk9, was required for the formation of a 5’–3’ gradient of H3K4 tri- to di- to monomethylation over genes (Wood et al., 2007; Xiao et al., 2007). Human PAF1 is also essential for H3K4 methylation (Kim et al., 2009; Zhu et al., 2005) and interacts closely with phosphorylated DSIF (Wier et al., 2013) to facilitate efficient elongation (Chen et al., 2009). It has also been hypothesized that H3K4 trimethylation occurs only after histone H3 substitution with variant H3.3 (Chen et al., 2013).

Where canonical H2A is suggested to be involved in gene silencing (Sauvageau and Sauvageau, 2010), histone variant H2A.Z potentially stimulates productive elongation (Venkatesh and Workman, 2015). Elongation by Pol II was 24% slower when Htz1, the yeast homolog to H2A.Z, is deleted in yeast—Htz1 was also found to displace nucleosomes during elongation and nucleosome occupancy increased in the transcribed regions of Htz1 mutant strains (Santisteban et al., 2011). H2A.Z was recently
shown to be enriched over a subset of MCF-7 cell enhancers and after induction with estrogen, these enhancers had stronger Pol II and Cohesin recruitment, more enhancer RNA transcription, and higher DNase I sensitivity than other enhancers (Brunelle et al., 2015). A study in Drosophila also reported that H2A.Z substitution facilitated Pol II transcription through early nucleosomes (Weber et al., 2014). H2A.Z is most frequently found with H3.3 and these double-variant nucleosomes as detected by sequential ChIP were particularly enriched near promoters in HeLa cells genome-wide—nucleosomes without this double substitution were relatively rare around the top 1,000-expressed genes (Jin et al., 2009). H2A.Z/H3.3 double-variants nucleosomes were also preferentially enriched for H3K4me3, further supporting a role in productive elongation (Yukawa et al., 2014). H2A.Z has also been described by some as a transcriptional repressor. Transcription by T7 RNA polymerase on an in vitro DNA template with a nucleosome containing H2A.Z/H3 or H2A.Z/H3.3, but not H2A/H3 or H2A/H3.3 was completely inhibited (Thakar et al., 2010). H2A.Z-containing nucleosomes were also more stable than canonical nucleosomes and blocked transcription by Pol II in HeLa nuclear extracts (Chen et al., 2013). H2A.Z/H3.3 nucleosomes, however, were more pliable than H2A.Z/H3 nucleosomes and were enriched at enhancers and gene bodies (Chen et al., 2013). More study is required to resolve these differences.

Enhancers are distinct genomic regions enriched in H3K4me1 and H3K27ac where Pol II and transcription factors accumulate and ultimately influence gene expression over great distances (Shlyueva et al., 2014) (Figure 3). Super-enhancers, or groups of active enhancers clustered over tens of kilobases, vary in factor occupancy between cell types, are highly associated with cell identity genes, and drive oncogene expression in a number of cancers (Hnisz et al., 2013). While there are about a million known enhancers (Andersson et al., 2014), their influence over genes is tightly constrained by incompletely understood mechanisms (Zabidi et al., 2015). Like promoters, enhancers can be active and relatively devoid of nucleosomes, poised and marked with a mixture of active and repressive histone modifications, or inactive and occluded by heterochromatin (Kim and Shiekhattar, 2015). Initiation at enhancers is bidirectional and very closely resembles promoters (Core et al., 2014) although...
Figure 3. Chromatin around promoters and enhancers
Histone modifications found near enhancers (top) and promoters (bottom) are illustrated and will be discussed in depth in Chapter 2. Transcription factors bound at enhancers influence nearby promoters through Mediator and Cohesin looping.

Enhancer RNAs are produced at equivalent levels in both directions unlike promoters which are generally sense-biased (Kim et al., 2010; Wang et al., 2011). It is unclear if enhancer transcripts are merely a consequence of Pol II engagement or if they serve functional purposes (Li et al., 2016). Synthesis of these transcripts was upregulated after estrogen signaling in MCF-7 cells (Hah et al., 2011) and more broadly as the earliest detected response in several mouse and human cell types to a number of stimuli (Arner et al., 2015). Mediator and Cohesin associate with Pol II at enhancers in a cell-type-specific manner and promote chromatin looping between enhancers and promoters (Kagey et al., 2010). This enhancer looping is thought to be responsible for enhancer-mediated transcriptional activation. Interestingly, Cohesin and Mediator were found at enhancers associated with both active and completely silent genes in embryonic cells, suggesting that enhancer looping might be primed, but not yet established at some loci (Lin et al., 2013). Broader CTCF looping is thought to restrict
enhancer-promoter looping to “insulated neighborhoods” (Dowen et al., 2014) and these topologically associating domains explain how enhancer transactivation is regulated (Ji et al., 2016).

**Focus of the Thesis**

The goal of the research presented here is to investigate the mechanisms which control Pol II transcription and co-transcriptional mRNA capping. In Chapter 2, sequencing methods were optimized and used to study the genome-wide occupancies of Pol II, pausing and elongation factors, and histone variants and modifications. I found that H3K4me3 is selectively associated with productive elongation downstream of genes, while H2A.Z is enriched tightly around promoters and more broadly in enhancer regions. In collaboration with Alberto Bosque and Vicente Planelles, I also discovered a role for transcription factor STAT5A in regulating HIV in a primary T cell model of latency and reactivation. That data was also used to reveal a new method of regulating *HEXIM1* gene expression. I also analyzed Myc-Max occupancy in vitro and in cells and demonstrated that Myc frequently associates with transcription machinery and not DNA sequence elements. In Chapter 3, a recently developed anti-cancer compound THZ1 was tested in vitro to establish its mechanism of action. I found that Cdk7, the covalent target of THZ1, and its activities during initiation were required for an ordered exchange of initiation factors with DSIF, NELF, and HCE. I also characterized the spatial and temporal requirements for co-transcriptional mRNA capping and uncovered the activity of a previously undescribed modulator of capping. In Chapter 4, in vitro and sequencing methods were used to describe an extremely rapid and global transcriptional response to hydrogen peroxide. By performing these experiments with or without flavopiridol, a potent P-TEFb inhibitor, I found evidence that supports a new role for P-TEFb in termination. In Chapter 5, the significant findings of this research are summarized and conclusions are integrated with recent results in the literature. I also discuss future development and use of a technique that combines isolation of nascent transcripts from cells, separation of methylated and unmethylated capped RNAs, and sequencing.
CHAPTER 2: GENOME-WIDE STUDIES OF ELONGATION CONTROL

This chapter focuses on the development and use of sequencing techniques to study transcription elongation control genome-wide. Portions of this chapter were previously published (Guo et al., 2014a; Liu et al., 2014; Nilson and Price, 2011) and are reused with permission.

Introduction

Control of the elongation phase of transcription by Pol II is essential in regulating gene expression (Guo and Price, 2013), but the mechanisms involved are incompletely understood. Evidence of promoter-proximal Pol II pausing has existed for over 30 years, but the prevalence of this pausing across the genomes of higher eukaryotes was not fully realized until sequencing experiments became widespread (Adelman and Lis, 2012). Through the use of chromatin immunoprecipitation (ChIP) followed by microarray, Pol II was discovered to accumulate near transcription start sites in human primary lung fibroblasts (Kim et al., 2005). While originally thought to be preinitiation complexes, later experiments showed these polymerases were accompanied by chromatin marks that were indicative of transcription initiation (Guenther et al., 2007). Pol II was later shown to initiate and elongate a short distance in both sense and antisense directions around most promoters (Core et al., 2008; Seila et al., 2008) and ChIP followed by sequencing (ChIP-Seq) was used to demonstrate that promoter-proximal pausing of Pol II by DSIF and NELF occurs on essentially all expressed genes in metazoans (Rahl et al., 2010). Thousands of transcription factor and chromatin mark ChIP-Seq datasets have since been produced in a number of cell types (Consortium, 2012). Promoter-proximal pausing at bidirectional promoters and enhancers is thought to maintain transcriptionally permissive chromatin environments and provide platforms for transcription factors (Core et al., 2014; Gilchrist et al., 2008; Scruggs et al., 2015), but questions remain about their regulation and purpose.

Transcription elongation control features prominently in the post-integration phases of human immunodeficiency virus (HIV) infection (Nilson and Price, 2011). Upon crossing the mucosa, HIV docks through interactions with cell surface proteins CD4 and CCR5 or CXCR4 and after fusion, releases viral single-stranded RNA, reverse
transcriptase, and integrase into the cytoplasm. The resulting double-stranded viral DNA is chaperoned into the nucleus where it integrates preferentially within the introns of highly expressed genes near the nuclear periphery (Cohn et al., 2015; Marini et al., 2015). An initial round of host-induced transcription results in expression of Tat, the primary transactivator of HIV, which directly interacts with P-TEFb (Tahirov et al., 2010), removes it from the 7SK snRNP (Barboric et al., 2007; Krueger et al., 2010; Sedore et al., 2007), and recruits it to the HIV promoter through an interaction with the structured viral nascent transcript (Jones and Peterlin, 1994; Wei et al., 1998). Viral gene expression is further amplified through transcription factor binding elements, including NFκB and Sp1 sites (Nilson and Price, 2011), and Tat-mediated recruitment of super elongation complex members (Chou et al., 2013; Gu et al., 2014; He et al., 2010; Lu et al., 2014; Schulze-Gahmen et al., 2014; Schulze-Gahmen et al., 2013; Sobhian et al., 2010). If untreated, HIV over time will result in immunodepletion and onset of AIDS.

Although several antiretroviral therapies exist to control the spread of HIV within infected individuals, these drugs have no effect on cells that host dormant viral genomes (Chun et al., 2015). Latent HIV reservoirs exist in a number of cell types including naïve T cells, macrophages, and others, but the most long-lived and readily reactivated reservoir is found in resting memory CD4-positive T cells (Churchill et al., 2016). Reasons for viral latency include the naturally low levels of P-TEFb (Budhiraja et al., 2013; Zhu et al., 1997) and NFκB (Pazin et al., 1996) in resting T cells, transcriptional interference from nearby genes (Greger et al., 1998; Han et al., 2008; Lenasi et al., 2008), and silencing by chromatin (Pearson et al., 2008; Tyagi et al., 2010). Because these cells have the ability to initiate new rounds of viral transcription at any time, infected patients must continue antiretroviral therapies for the duration of their natural lives (Nilson and Price, 2011). Genome-wide studies of HIV have been limited to model systems as it is not yet possible to separate latently infected cells from uninfected cells. Research into the mechanisms of HIV latency and reactivation could lead to the development of new drugs that force viral reactivation, which if administered alongside existing antiretroviral therapies, would prevent viral spread and lead to eradication of latent reservoirs (a strategy known as “shock and kill”). The diversity of latently infected cell types, however, could limit the effectiveness of this approach.
c-Myc is a transcription factor whose misregulation is associated with cancer (Dang, 2012; Luscher and Vervoorts, 2012). Myc has been shown to associate with P-TEFb and cause an increase in productive elongation at target genes (Eberhardy and Farnham, 2002; Gargano et al., 2007; Kanazawa et al., 2003). Myc also increases transcript cap methylation, which increases mRNA stability and stimulates translation (Cole and Cowling, 2009; Cowling and Cole, 2010). Myc forms a heterodimer with its binding partner Max (Blackwood and Eisenman, 1991; Blackwood et al., 1992; Nair and Burley, 2003) and this complex has been demonstrated to bind preferentially to the palindromic “E-box” sequence CACGTG and related variants (Blackwell et al., 1993; Blackwell et al., 1990). These initial observations and a number of studies on the effects of Myc on specific genes led to a prevailing model where Myc, in conjunction with Max, binds to E-box sequences and subsequently regulates Pol II transcription (Dang, 2012; Eilers and Eisenman, 2008; Luscher and Vervoorts, 2012; Meyer and Penn, 2008). This model does not adequately explain why the genes regulated by Myc change depending on the cell types and conditions used (Littlewood et al., 2012). Two recent studies used inducible systems to show that Myc, when switched on, amplified transcription of genes already being expressed (Lin et al., 2012a; Nie et al., 2012). Myc has since been shown to modulate global gene expression by associating with super-enhancers (Chipumuro et al., 2014; Christensen et al., 2014; Loven et al., 2013). Deciphering the relative contributions of protein-DNA and protein-protein interactions in Myc recruitment could explain the global effects of Myc and other transcription factors.

Materials and Methods

ChIP-Seq

The following antibodies were used for chromatin immunoprecipitation: Pol II (Santa Cruz Biotechnology sc-899), NELF (NELF-A) (SCBT sc-23599), human capping enzyme (HCE) (Adamson et al., 2005), DSIF (SPT5) (SCBT sc-28678), H3K4me1 (Abcam ab8895), H3K4me3 (Abcam ab8580), H2A.Z (Millipore 07-594), and STAT5A (SCBT sc-1081).

For ChIP-Seq in suspension HeLa cells, HeLa cells were grown to $5 \times 10^5$ cells/ml in spinner flasks at 37°C and 5% CO$_2$ in SMEM (Gibco 11380-037) supplemented with 2 mM L-Glutamine (Gibco 25030-081) and 10% FBS (Gibco 26140-079). Cells were removed
from the incubator and crosslinked promptly by adding 16% paraformaldehyde (Electron Microscopy Sciences 130218) to a final concentration of 1% in media. Crosslinking was allowed to proceed for 10 min (Pol II, DSIF, and NELF), or 15 min (H3K4me1, H3K4me3, and H2A.Z) before glycin addition to 125 mM (Pol II, DSIF, and NELF) or Tris pH 7.6 addition to 1.33 M (H3K4me1, H3K4me3, and H2A.Z). All solutions used for steps after crosslinking and before elution were ice cold and supplemented with EDTA-free protease inhibitor cocktail (Roche 11873580001), 0.1% isopropanol-saturated PMSF, and 1 mM DTT. Cells were pelleted at 1200 × g for 5 min at 4° and rinsed twice with PBS. Decanted, rinsed pellets were stored at -80°.

For Pol II, DSIF, and NELF ChIP-Seq experiments, 5 × 10⁷ cells were used per ChIP. Pellets were thawed and incubated for 10 min on ice in hypotonic buffer (20 mM HEPES pH 7.9, 10 mM KCl, 1 mM EDTA, and 10% glycerol), Dounce homogenized for 30 strokes, pelleted at 16,000 × g for 3 min at 4°, washed once in hypotonic buffer, incubated for 30 min on ice with occasional vortexing in RIPA buffer (50 mM Tris pH 7.6, 150 mM NaCl, 1 mM EDTA, 0.25% sodium deoxycholate, and 1% IGEPAL CA-630), and sonicated in 5 × 10⁷ cell per 1 ml RIPA buffer aliquots on wet ice using a Fisher Model 550 Sonic Dismembrator (Fisher Scientific) for 21 cycles of 30 s at intensity 4, followed by 90 s off. After sonication, samples were pelleted at 20,000 × g for 15 min at 4° and supernatants were stored at -80°. 1% of each sonicated sample was set aside for input sequencing. Sonicated samples were incubated with 10 μg antibody overnight at 4° with rotation, and then with 50 μl Protein G Sepharose 4B Fast Flow bead slurries (Sigma P3296) (equilibrated in RIPA buffer) for 1 h at 4° with rotation. Beads were transferred to Bio-Spin disposable columns (Bio-Rad 732-6008) and washed with 15 ml RIPA buffer and 10 ml PBS. Beads were eluted in elution buffer (20 mM Tris pH 7.6, 1 mM EDTA, and 1% SDS), incubated overnight at 65° to reverse crosslinks, and treated with 2 μl RNase A (Thermo EN0531) for 30 min at 37°, then 2 μl Proteinase K (Thermo EO0491) for 30 min at 50°. DNA was isolated by MinElute PCR Purification Kit (QIAGEN 28004) and libraries were prepared and sequenced by the Iowa Institute of Human Genetics Genomics Division on an Illumina HiSeq 2000 using 100 bp paired-end reads.

For H3K4me1, H3K4me3, and H2A.Z ChIP-Seq experiments, 2.5 × 10⁷ cells were used per ChIP. Pellets were thawed in 1 ml 2% sarkosyl buffer (10 mM Tris pH 7.6, 100
mM NaCl, 1 mM EDTA, and 2% sarkosyl). This buffer would normally contain 0.2% sarkosyl but was mistakenly made with 2% sarkosyl. Thawed cells were homogenized with 10 loose pestle strokes in a 1 ml Dounce, pelleted at 16,000 × g for 3 min at 4°, washed once with 1 ml 2% sarkosyl buffer, and sonicated in 14 ml polystyrene tubes on wet ice with a Fisher Model 550 Sonic Dismembrator for 36 cycles of 30 s at intensity 4, followed by 90 s off. After sonication, samples were pelleted at 16,000 × g for 15 min at 4° and supernatants were stored at -80°. Sonicated samples were precleared with 32.5 μl each of Protein A (Sigma P9424) and Protein G Sepharose 4B Fast Flow bead slurries (equilibrated in 2% sarkosyl buffer) for 2 h at 4° with rotation. After settling, precleared sonicated samples were transferred to new tubes and incubated with 10 μg antibody overnight at 4° with rotation. 1% of each precleared sample was set aside for input sequencing. In parallel, 32.5 μl each of Protein A- and Protein G Sepharose 4B Fast Flow bead slurries were washed with 1 ml 2% sarkosyl buffer and blocked in 1 ml 2% sarkosyl buffer containing 1 mg/ml BSA (New England BioLabs B9000S) overnight at 4° with rotation. Blocked beads were washed with 1 ml 2% sarkosyl buffer and incubated with chromatin-antibody samples for 2 h at 4° with rotation. Beads were settled and washed four times with 1 ml wash buffer (10 mM Tris pH 7.6, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and 0.1% SDS) and two times with 1 ml rinse buffer (10 mM Tris pH 7.6, 50 mM NaCl, and 1 mM EDTA). Beads were eluted in 100 μl elution buffer 1 (10 mM Tris pH 7.6, 1 mM EDTA, and 1% SDS) for 10 min at 65°, and 150 μl elution buffer 2 (10 mM Tris pH 7.6, 1 mM EDTA, and 0.67% SDS) for 10 min at 65°. Pooled eluates were treated with 2 μl RNase A for 30 min at 37° and then incubated with 2 μl Proteinase K for 2 h at 65° to reverse crosslinks. DNA was isolated by MinElute PCR Purification Kit and libraries were prepared using a NEXTflex Rapid DNA-Seq Kit (Bioo Scientific 5144-04). Sequencing was performed by the Iowa Institute of Human Genetics Genomics Division on an Illumina HiSeq 2500 using 125 bp paired-end reads.

For ChIP-Seq experiments in a primary T cell model of HIV latency and infection, cell culture and treatments were performed by Alberto Bosque at the University of Utah. Central-memory-like T cells were generated from peripheral blood mononuclear cells obtained using Leukopaks from unidentified, healthy donors and infected as described previously (Bosque and Planelles, 2009, 2011; Martins et al., 2016).
For Pol II ChIP-Seq in uninfected or HIV-infected cells ± αCD3/αCD28, quiescent cells were left untreated or incubated for 24 h with 1 bead/cell of Human T-Activator CD3/CD28 for T Cell Expansion and Activation Dynabeads (Life Technologies 11131D). For Pol II and STAT5A ChIP-Seq in HIV-infected cells ± HODHBt, quiescent cells were left untreated or incubated for 3 h with IL-2 ± 100 μM HODHBt (AK Scientific P962). Cells were removed from the incubator and crosslinked promptly by adding 16% paraformaldehyde to a final concentration of 1% in media. Crosslinking was allowed to proceed for 2 min (Pol II in uninfected or HIV-infected primary T cells ± αCD3/αCD28) or 15 min (Pol II and STAT5A in HIV-infected primary T cells ± HODHBt) before Tris pH 7.6 addition to 1.33 M. All solutions used for steps after crosslinking and before elution were ice cold and supplemented with EDTA-free protease inhibitor cocktail and 0.1% isopropanol-saturated PMSF. Cells were pelleted at 1000 × g for 5 min at 4° and rinsed twice with PBS. Decanted, rinsed pellets were stored at -80°.

For Pol II ChIP-Seq in uninfected or HIV-infected cells ± αCD3/αCD28, 4-6 × 10⁶ cells were used per ChIP. Pellets were thawed in 1 ml 0.2% sarkosyl buffer (20 mM Tris pH 7.6, 150 mM NaCl, 1 mM EDTA, and 0.2% sarkosyl), homogenized with 10 loose pestle strokes in a 1 ml Dounce, and sonicated in 14 ml polystyrene tubes on wet ice with a Fisher Model 550 Sonic Dismembrator for 20 cycles of 30 s at intensity 2.5, followed by 90 s off. After sonication, samples were pelleted at 16,000 × g for 15 min at 4° and supernatants were stored at -80°. Sonicated samples were precleared with 32.5 μl each of Protein A and Protein G Sepharose 4B Fast Flow bead slurries (equilibrated in 0.2% sarkosyl buffer) for 1 h at 4° with rotation. After settling, precleared sonicated samples were transferred to new tubes and incubated with 10 μg antibody overnight at 4° with rotation. 1% of each precleared sample was set aside for input sequencing. In parallel, 32.5 μl each of Protein A and Protein G Sepharose 4B Fast Flow bead slurries were washed with 1 ml 0.2% sarkosyl buffer and blocked in 1 ml 0.2% sarkosyl buffer containing 1 mg/ml BSA overnight at 4° with rotation. Blocked beads were washed with 1 ml 0.2% sarkosyl buffer and incubated with chromatin-antibody samples for 2 h at 4° with rotation. Beads were transferred to Bio-Spin disposable columns (blocked first with 1 ml 0.2% sarkosyl buffer containing 1 mg/ml BSA) and washed with 4 ml wash buffer and 1 ml rinse buffer. Beads were eluted in 100 μl elution buffer 1 for 10 min at
65°, and 150 μl elution buffer 2 for 10 min at 65°. Pooled eluates were treated with 2 μl RNase A for 30 min at 37° and then incubated with 2 μl Proteinase K for 2 h at 65° to reverse crosslinks. DNA was isolated by MinElute PCR Purification Kit and libraries were prepared using a NEXTflex Rapid DNA-Seq Kit. Sequencing was performed by the UCLA Neuroscience Genomics Core on an Illumina HiSeq 2500 using 69 bp and 107 bp paired-end reads.

For Pol II and STAT5A ChIP-Seq in HIV-infected cells ± HODHBt, 1.2 × 10⁶ cells were used per ChIP. Pellets were thawed in 1 ml 0.2% sarkosyl buffer, homogenized with 10 loose pestle strokes in a 1 ml Dounce, and sonicated in 14 ml polystyrene tubes on wet ice with a Fisher Model 550 Sonic Dismembrator for 35 cycles of 30 s at intensity 4, followed by 60 s off. After sonication, 100 μl 10% Triton X-100 was added and samples were pelleted at 16,000 × g for 15 min at 4°. Supernatants were stored at -80°. Sonicated samples were precleared with 50 μl PBS-washed Protein A Dynabeads (Life Technologies 10001D) for 3 h at 4° with rotation. In parallel, 50 μl Protein A Dynabeads were PBS-washed, blocked in 50 μl PBS containing 1 mg/ml BSA for 1 h at 4° with rotation, washed again with PBS, and incubated with 15 μg antibody for 2 h at 4° with rotation. Antibody-beads were washed twice with 100 μl wash buffer A (20 mM Tris pH 7.6, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and 0.1% SDS), then resuspended with precleared sample and incubated overnight at 4° with rotation. 1% of each precleared sample was set aside for input sequencing. Samples were washed once with 200 μl wash buffer A, 200 μl wash buffer B (20 mM Tris pH 7.6, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and 0.1% SDS), and 200 μl wash buffer C (20 mM Tris pH 7.6, 250 mM LiCl, 1 mM EDTA, 0.5% IGEPA CA-630, 0.5% sodium deoxycholate, and 0.1% SDS), and twice with 200 μl rinse buffer. Samples were incubated with 200 μl elution buffer for 2 h at 65° to reverse crosslinks. Eluates were treated with 2 μl RNase A for 30 min at 37°, then 2 μl Proteinase K for 30 min at 50°. DNA was isolated by MinElute PCR Purification Kit and libraries were prepared using a NEXTflex Rapid DNA-Seq Kit. Sequencing was performed by the Iowa Institute of Human Genetics Genomics Division on an Illumina HiSeq 2500 using 125 bp paired-end reads.
For all ChIP-Seq experiments, raw paired sequences were first aligned to the UCSC hg19 assembly using Bowtie 2.2.6 (--no-mixed --no-discordant --reorder) (Langmead and Salzberg, 2012). Human-mapped reads were compiled using MACS 2.0.10 (--format BAMPE --bdg) (Zhang et al., 2008), compressed (Kent et al., 2010), and visualized on the UCSC Genome Browser (Kent et al., 2002). For HIV ChIP-Seq experiments, paired sequences where both mates failed to map to hg19 (SAM flags 77 or 141) were then aligned to the NL4-3 HIV-1 genome using Bowtie 1.1.1 (-v 3 -m 2 --all --offbase 1 --maxins 2000) (Langmead et al., 2009) and unaligned sequences were saved (-un). Unmapped reads whose mate mapped to hg19 (SAM flags 69 or 133) were also aligned to the HIV genome in an attempt to identify fragments which spanned integration sites. HIV alignments were also performed where sequences were trimmed from the 5′ end to 36 bp in an attempt to identify reads which spanned integration sites. Human integration sites were confirmed using BLASTn. ChIP-Seq tracks were compiled from all HIV-mapping fragments and after changing the chromosome field to “chr1” were visualized using a track hub (Raney et al., 2014) on the UCSC Genome Browser. Fragments which mapped twice, once to each LTR, were color-coded and both copies were retained. ChIP-Seq tracks were normalized as indicated in figure legends by dividing the number of hg19-mapped reads within each sample by the average number of hg19-mapped reads across samples within an experiment.

Metagene analyses were performed as described previously (Cheng et al., 2012) with the following modifications. A new gene list was generated for the hg19 genome assembly. To do this, the hg19 RefSeq refGene database (50,062 entries) was downloaded and redundant entries, defined as genes with transcription start sites (TSS) within 100 bp of the next closest TSS on the same strand, were compressed into one entry by averaging their txStart values. To prevent misinterpretation of signals from annotated divergent promoters, genes whose TSS were still within 1 kb of the next closest TSS on any strand were removed. Finally, all miRNA, U6 RNA, and snoRNA genes were removed, resulting in a gene list with 20,868 entries. The background subtraction step was also changed to use the single lowest point instead of the average of the lowest 2,000 values in each curve. The areas under all curves ± 10 kb were normalized to 1.
Heatmaps were generated using the computeMatrix and plotHeatmap functions of deepTools 2.2.4 (--referencePoint=TSS --sortUsing=median) (Ramirez et al., 2014) and the same gene list used for metagene analyses.

**RNA-Seq**

Cell culture and treatments were performed by Alberto Bosque at the University of Utah. Central-memory-like T cells were generated from peripheral blood mononuclear cells obtained using Leukopaks from three unidentified, healthy donors as described previously (Bosque and Planelles, 2009, 2011; Martins et al., 2016). Quiescent cells were incubated for 24 h with IL-2 alone (control), with IL-2 and 100 μM HODHBt, or with Human T-Activator CD3/CD28 for T Cell Expansion and Activation Dynabeads. Total RNA was extracted using Trizol (Invitrogen 15596), polyadenylated RNA was enriched using NEXTflex Poly(A) Beads (Bioo Scientific 512980), and stranded libraries were prepared using a NEXTflex Rapid Directional qRNA-Seq Kit (Bioo Scientific 5130-02D). Sequencing was performed by the Iowa Institute of Human Genetics Genomics Division on an Illumina HiSeq 2500 using 125 bp paired-end reads.

Raw paired sequences were first trimmed to remove Illumina sequencing adapters and Bioo Scientific randomized 8-mer molecular indexes using Trim Galore! 0.4.1 (--quality 0 --illumina --paired --clip_R1 9 --clip_R2 9 --three_prime_clipR1 9 --three_prime_clipR2 9), and then aligned to the UCSC hg19 genome assembly using TopHat 2.1.0 (--library-type fr-firststrand --no-discordant --no-mixed) (Kim et al., 2013). A custom script was then used to remove PCR duplicates by searching for identical fragments with identical randomized 8-mer molecular indexes. For RNA-Seq tracks, reads were then compiled using the genomecov (-split) function of bedtools 2.25.0, normalized by dividing the number of hg19-mapped reads within each sample by the average number of hg19-mapped reads across all nine samples, combined across donors, and visualized on the UCSC Genome Browser. For differential expression analysis, a pipeline developed by Alec McIntosh and Miles Pufall was used. Reads were counted over UCSC hg19 genes (26,364 entries) using the featureCounts (-p -B -C -M) function of Subread 1.5.0 (Liao et al., 2014). These feature counts were imported into R and normalized using both EDASeq 2.4.1 (Risso et al., 2011), which reduces variability from different sequencing depths between replicates, and RUVSeq 1.4.0 (Risso et al.,
which when given a list of all genes to use for normalization (RUVr method) further reduces variability from batch effects in an unbiased manner. Differential expression analysis was then performed using DESeq2 1.10.1 (test = “Wald”, fitType = “parametric”, betaPrior = TRUE, and pAdjustMethod = “fdr”) (Love et al., 2014) using settings which prevented additional normalization.

Myc occupancy analyses

Protein binding microarray conditions were previously published (Guo et al., 2014a). Statistical analyses were performed in R. In Figure 21 the frequencies of CACGTG sites in Myc-occupied and DNase I-hypersensitive regions were compared using Fisher’s exact test (dhyper). A similar analysis was performed using Pol II ChIP-Seq peaks instead of CACGTG sequences. In Figure 22, in vitro Myc 8-mer occupancies were compared using the Wilcoxon rank-sum test (wilcox.test with the paired parameter set to false). A receiver operating characteristic (ROC) curve was used to assess the enrichment of sites with high in vitro 8-mer occupancies in the Myc ChIP-Seq peaks. A ROC curve is a plot of false positive rate (1-specificity) versus true positive rate (sensitivity) computed for different 8-mer occupancy cutoffs. Each point on the ROC curve corresponds to one cutoff. Sequences with 8-mer occupancy above or below the cutoff are predicted positives (that is, predicted to be bound by Myc in vivo) or predicted negatives (that is, predicted not bound by Myc in vivo), respectively. In Figure 23A, the correlation between Myc ChIP-Seq and in vitro occupancies was tested using Spearman’s rank correlation coefficient (cor.test with the method parameter set to Spearman).

Nuclear walk-on

Nuclei isolation conditions were adapted from (Chao and Price, 2001). HeLa cells were grown to $5 \times 10^5$ cells/ml in spinner flasks at 37° and 5% CO$_2$ in SMEM supplemented with 10% FBS and 2 mM L-Glutamine. Suspension cells were pelleted at 500 × g for 10 min, media was decanted, and pellets were resuspended in 15 ml ice-cold lysis buffer (10 mM Tris pH 7.6, 320 mM sucrose, 0.5% Triton X-100, 2 mM Mg(C$_2$H$_3$O$_2$)$_2$, 3 mM CaCl$_2$, 1 mM DTT, 0.004 U/μl SUPERase-In (Ambion AM2696), and 0.1% isopropanol-saturated PMSF) and Dounce homogenized. Cell lysis was monitored by
phase-contrast microscopy. Once lysed, cells were pelleted at 1,200 × g for 5 min and thoroughly resuspended in a mixture of 1 ml lysis buffer and 2 ml sucrose cushion (10 mM Tris pH 7.6, 1.9 M sucrose, 5 mM Mg(C₂H₃O₂)₂, 1 mM DTT, 0.004 U/μl SUPERase-In, and 0.1% isopropanol-saturated PMSF). This homogenate was carefully layered over a 2 ml sucrose cushion and spun at 30,000 × g for 45 min. Nuclei were resuspended in 1 ml storage buffer (10 mM Tris pH 7.6, 25% glycerol, 5 mM Mg(C₂H₃O₂)₂, and 5 mM DTT), pelleted at 1,200 × g for 5 min, and resuspended with storage buffer to about 5 × 10⁷ nuclei/ml before storage in aliquots at -80°.

6 × 10⁶ nuclei were diluted to 120 μl in storage buffer lacking glycerol. Five 20 μl aliquots were then incubated for 3 min at 30° with an equal volume of reaction buffer containing 10 mM Tris pH 7.6, 1% sarkosyl, 300 mM KC₂H₃O₂, 5 mM Mg(C₂H₃O₂)₂, 5 mM DTT, 0.5 U/μl SUPERase-In, and one of the following nucleotide concentrations: 0.33 μM α-³²P-CTP (PerkinElmer BLU008H001MC) only; 10 μM biotin-11-ATP (PerkinElmer NEL544), 10 μM biotin-11-UTP (PerkinElmer NEL543), 10 μM biotin-11-GTP (PerkinElmer NEL545), and 0.33 μM α-³²P-CTP; 10 μM ATP (Thermo R0441), 10 μM 5-bromouridine 5'-triphosphate (BrUTP) (Sigma B7166), 10 μM GTP (Thermo R0461), and 0.33 μM α-³²P-CTP; or 250 μM ATP, 250 μM BrUTP, 250 μM GTP, 1 μM CTP (Thermo R0451), and 0.33 μM α-³²P-CTP. Reactions were stopped by addition of EDTA to 20 mM and labeled transcripts were extracted with Trizol LS (Ambion 10296-028) and precipitated with 95% ethanol and 500 mM NH₄C₂H₃O₂. Further treatments of nascent transcripts are described below. To visualize results, RNAs were separated on denaturing RNA gels (6 M urea, 1X TBE, and 6 or 9% 37.5:1 acrylamide:bis-acrylamide), scanned using a Fujifilm Typhoon FLA-7000 phosphorimager, and analyzed using Fujifilm MultiGauge v3 software.

**In vitro transcription**

The use of soluble and paramagnetic bead-immobilized CMV templates and HeLa nuclear extract (HNE) to study transcription in vitro was described previously (Adamson et al., 2003; Cheng and Price, 2007, 2009). All steps were performed at room temperature (RT). Template DNA from CMV -800 bp to +508 bp was incubated for 30 min with 1 μl/rxn HNE in the presence of 60 mM KCl, 5 mM MgCl₂, 20 mM HEPES pH 7.6, 1 mM DTT, and 0.5 U/μl SUPERase-In. Initiation was accomplished with a 30 s pulse
containing 60 mM KCl, 5 mM MgCl₂, 20 mM HEPES pH 7.6, 0.21 μM α³²P-CTP, and 500 μM ATP/UTP/GTP (limiting CTP). Elongation complexes were either stopped by addition of EDTA to 20 mM or isolated with high salt wash (1.6 M KCl, 20 mM HEPES pH 7.6, 1 mM DTT, and 0.02% Tween20) to remove associated factors. Labeled transcripts were extracted with phenol and precipitated with 95% ethanol and 500 mM NH₄C₂H₃O₂. Further treatments of in vitro transcripts are described below.

**Transcript separation techniques**

Dynabeads M-280 Streptavidin (Invitrogen 11206D) were washed twice in 2 M NaCl buffer (20 mM Tris pH 7.6, 2 M NaCl, 1 mM EDTA, and 0.02 U/μl SUPERase-In), incubated twice for 2 min in 100 mM NaOH and 50 mM NaCl, and washed two more times in 2 M NaCl buffer. Precipitated transcripts were resuspended in a buffer containing 25 mM Tris pH 7.6 and 0.5 U/μl SUPERase-In and incubated with beads for 1 h at RT with rotation. Samples were then washed once with 50 μl Dynabead high salt wash buffer (20 mM Tris pH 7.6, 2 M NaCl, and 0.5% Triton X-100) and once with 50 μl Dynabead medium salt wash buffer (20 mM Tris pH 7.6, 300 mM NaCl, and 0.1% Triton X-100); unbound supernatants were saved and pooled. Bound and unbound fractions were brought to equal volumes in Dynabead medium salt wash buffer, spiked with glycogen, and Trizol LS extracted, precipitated, and analyzed as described above.

Anti-BrdU mouse mAb agarose beads (SCBT sc-32323 AC) were washed with bead binding buffer (20 mM Tris pH 7.6, 50 mM KC₂H₃O₂, 1 mM EDTA, 0.02% Tween20, and 0.02 U/μl SUPERase-In) and blocked in bead binding buffer containing 0.5 mg/ml BSA for 10 min at RT with rotation. Beads were then washed once with RIPA buffer (20 mM Tris pH 7.6, 150 mM KC₂H₃O₂, 1% Triton X-100, 0.2% sodium deoxycholate, 0.1% SDS, and 1 mM EDTA), once with LiCl buffer (20 mM Tris pH 7.6, 400 mM LiCl, 0.02% Tween20, and 1 mM EDTA), and three times with bead binding buffer. Precipitated transcripts were resuspended in a buffer containing 25 mM Tris pH 7.6 and 0.5 U/μl SUPERase-In and incubated with beads for 1 h at RT with rotation. Samples were then washed three times with 70 μl bead wash buffer (20 mM Tris pH 7.6, 150 mM KC₂H₃O₂, 1 mM EDTA, 0.1% Tween20, and 0.02 U/μl SUPERase-In); unbound supernatants were saved and pooled. Bound and unbound fractions were brought to equal volumes in bead
wash buffer, spiked with glycogen, and phenol extracted, precipitated, and analyzed as described above.

Cap status determination using recombinant human cap methyltransferase (HCM) was described previously (Moteki and Price, 2002; Nilson et al., 2015) and in greater detail in Chapter 3. 10 μg/rxn of Mono Q-purified anti-2,2,7-trimethylguanosine antibody was bound to 13 μl/rxn Protein G Sepharose 4B Fast Flow beads (equilibrated in bead binding buffer containing 0.5 mg/ml BSA) for 1 h at 4° with rotation. Antibody-beads were washed once with RIPA buffer, once with LiCl buffer, and three times with bead binding buffer. In parallel, precipitated transcripts were resuspended in a buffer containing 25 mM Tris pH 7.6 and 0.5 U/μl SUPERase-In and incubated for 30 min at 37° with an equal volume of 20 mM Tris pH 7.6, 10 mM Mg(C₂H₃O₂)₂, 200 μM SAM (NEB B9003S), 1 mM DTT, 0.2 mg/ml BSA, and 50 ng/μl HCM to methylate existing RNA caps. After methylation, transcripts and antibody-beads were incubated together with rotation for 2 h at RT. Samples were then washed three times with 70 μl bead wash buffer; unbound supernatants were saved and pooled. Bound and unbound fractions were brought to equal volumes in bead wash buffer, spiked with glycogen, and phenol extracted, precipitated, and analyzed as described above.

CIP and HCE add-backs

Precipitated transcripts were resuspended in reaction buffer (50 mM Tris pH 7.6, 100 mM K₂H₃O₂, 10 mM Mg(C₂H₃O₂)₂, 1 mM DTT, and 0.02 U/μl SUPERase-In) and divided into three equal portions. Mock samples were incubated with an equal volume of reaction buffer for 30 min at 37°. CIP samples were incubated with an equal volume of reaction buffer containing 2.5 U/μl Calf Intestinal Alkaline Phosphatase (CIP) (NEB M0290S) for 30 min at 37°. HCE samples were incubated with an equal volume of 20 mM Tris pH 7.6, 10 mM Mg(C₂H₃O₂)₂, 1 mM DTT, 0.4 mg/ml BSA, 0.1 mg/ml recombinant human capping enzyme (HCE) (Moteki and Price, 2002), and 2 mM GTP for 30 min at 37°. Transcripts were then methylated using HCM and SAM, separated using anti-2,2,7-trimethylguanosine antibody, spiked with glycogen, phenol extracted, precipitated, and analyzed as described above.
Results

Optimization of chromatin immunoprecipitation

Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) enables identification of the genome-wide occupancy of chromatin-associated factors. ChIP-Seq has four overall steps: crosslinking, sonication, immunoprecipitation, and sequencing (Figure 4). In the first step, protein-protein and protein-DNA associations present within cells are covalently bridged by chemical agents (generally 1% formaldehyde). Crosslinked cells are then pelleted, washed with PBS, lysed, and sonicated, which shears crosslinked chromatin into fragments appropriately sized for sequencing (~300 bp). Antibodies against target proteins are coupled with magnetic or agarose beads and incubated with samples, which are then washed to reduce non-specific binding. After immunoprecipitation, proteins are digested with Proteinase K, crosslinks are reversed with heat (65°), and the now naked DNA is isolated using either phenol-chloroform extraction or a silica-membrane-based DNA purification kit. The resulting DNA fragments are then prepared for sequencing. Because these fragments were selected by their association with a target protein, the sum of aligned fragments over a given genomic location is representative of that protein’s occupancy. Altering the conditions of any of these steps influences the overall results. For example, increasing the detergent enables more efficient sonication, but can be detrimental to antibody binding. Likewise, some cell types are more susceptible to formaldehyde crosslinking than others. ChIP-Seq conditions should be optimized for every experiment and tested before committing to sequencing.

Many ChIP-Seq protocols include a step in which formaldehyde is quenched by timed addition of 125 mM glycine (Landt et al., 2012) as excessive crosslinking hinders sonication and immunoprecipitation. At peak efficiency, formaldehyde forms a single methylene link between two proteins—an incomplete reaction will add a reactive methylol group to a protein (Travers and Buckle, 2000). Because 1% formaldehyde is 333 mM, 125 mM glycine can at best only quench a third of any potentially reactive products in the cell. When I increased the glycine quench to 750 mM, however, I saw no differences in sonication efficiency over a crosslinking time-course between 30 seconds and 10 minutes.
To address the possibility that glycine does not quench formaldehyde, I compared the sonication efficiencies of cells crosslinked normally or with formaldehyde pre-treated with glycine. First, I crosslinked suspension HeLa cells directly in 37° media by addition of methanol-free paraformaldehyde (which depolymerizes to formaldehyde in solution) to a 1% final concentration and, after 10 minutes, introduced glycine to 750 mM. Because suspension cells must be pelleted
before formaldehyde-containing media can be decanted, unquenched formaldehyde can act for another 5-10 minutes. After pelleting and PBS washing, I lysed and sonicated the crosslinked cells for 20 minutes using a Covaris S2 instrument, which is capable of consistent shearing between samples. Sheared chromatin was then divided equally and half underwent normal crosslink reversal with Proteinase K and overnight 65° incubation. This crosslink-reversed DNA represents the sonication efficiency over the entire DNA sample. Crosslink reversal was skipped on the other half and after phenol-chloroform isolation, only DNA which was never crosslinked to proteins will be recovered. This method to select uncrosslinked sonicated DNA can be used to isolate regions of naked chromatin (Giresi et al., 2007).

After 10 minutes of crosslinking before quenching with glycine, a range of fragment sizes was observed which spanned the entire gel (about 100-20,000 bp); this is indicative of poor sonication (Figure 5, lane 1). Very little DNA was observed when crosslink reversal was skipped, which suggests that almost all fragments were crosslinked by these conditions (Figure 5, lane 2). Next, I preincubated paraformaldehyde with a 3-fold molar excess of glycine for 3 minutes and this mixture was added to cells for 30 seconds before otherwise identical pelleting, washing, lysis, sonication, and isolation steps. This condition resulted in the same pattern of sheared DNA (Figure 5, lane 3) and almost all fragments were crosslinked (Figure 5, lane 4). These results strongly suggest that glycine does not effectively quench formaldehyde.

Next, I preincubated paraformaldehyde with Tris, which is used to quench formaldehyde in mass spectrometric experiments (Wu et al., 2011). Unlike the prequench with glycine, chromatin from cells treated with Tris-prequenched paraformaldehyde was sheared to a size range of 100-1500 bp (Figure 5, lane 5). Additionally, a nearly identical amount and distribution of DNA was observed when crosslink reversal was skipped (Figure 5, lane 6). These results show that Tris preincubation stops almost all formaldehyde crosslinking.

To determine if the amount of crosslinking could be controlled by timed addition of Tris, I tested both 5 and 1 minute crosslinking durations. Sonication was more efficient after 5 minutes of crosslinking (Figure 5, lane 7) when compared with the previous 10 minute condition (Figure 5, lane 1), but the target size of 300 bp was not
Figure 5. Tris, not glycine, quenches formaldehyde
Suspension HeLa cells were crosslinked by direct addition of methanol-free paraformaldehyde to a final concentration of 1% (333 mM) for indicated times. Crosslinking was quenched by addition of excess glycine or Tris pH 7.6. For lanes 3-6, paraformaldehyde was preincubated with glycine or Tris for 3 min before addition to cells for 30 s. Cells were then pelleted, lysed, and sonicated. The resulting fragmented chromatin was incubated with RNase A for 30 min at 37°C. Before phenol-chloroform extraction, + reversal samples were also incubated with Proteinase K for 30 min at 50°C; - reversal lanes were isolated immediately after RNase A treatment. 1% agarose and ethidium bromide. Achieved. Chromatin from cells crosslinked for 1 minute before Tris quench, however, was efficiently sheared to a size range suitable for sequencing (Figure 5, lane 9). Despite the short amount of crosslinking time, very few fragments were devoid of a protein-DNA crosslink (Figure 5, lane 10). Overall, these results demonstrate that the extent of formaldehyde crosslinking can be controlled by timed addition of Tris, but not glycine.

While Pol II exhibited excellent ChIP-Seq signal when isolated from cells crosslinked for only 1 minute with 1% formaldehyde, this low amount of crosslinking was insufficient to successfully determine the genomic occupancies of factors that loosely or transiently associate with chromatin (data not shown). Because other labs typically crosslink for 15-20 minutes (10 minutes plus pelleting time) and routinely immunoprecipitate proteins that do not bind DNA directly, I reasoned that longer
crosslinking times might improve ChIP of proteins associated with elongation complexes. Proteins with fewer accessible crosslinkable residues would also require additional crosslinking time to capture. Because formaldehyde only acts over 2 Å, other labs have experimented with alternative crosslinkers with longer spacer arms between their reactive groups. Glutaraldehyde, which has two aldehydes separated by 5 Å, was used at 1% for 10 minutes to successfully ChIP both Cdk9 and HEXIM1 (Ji et al., 2013). Ethylene glycol bis(succinimidyl succinate) (EGS), which has a 16.1 Å spacer arm, can also be used alongside formaldehyde for ChIP assays (Bittencourt et al., 2012; Zeng et al., 2006) or to capture distant chromatin interactions (Zhang et al., 2012b). To determine if additional crosslinking would improve the ChIP signals of indirect factors, I performed ChIP-qPCR against Pol II, NELF, and human capping enzyme (HCE) using cells crosslinked first with 1% paraformaldehyde for 1 or 20 minutes, and then with either 1.5 mM EGS for 15 minutes or 0.1% glutaraldehyde for 5 minutes.

ChIP-qPCR was performed using primers which amplified regions around Myc known by ChIP-Seq to be unoccupied by Pol II (-3823 bp upstream), or possess a promoter-proximal paused polymerase (+83 bp downstream). Fold enrichment over background was determined by dividing ChIP/input signals from these promoter-proximal and unoccupied regions. As expected, strong Pol II enrichment was detected using 1 or 20 minutes of paraformaldehyde (Figure 6). NELF ChIP signal-to-noise improved nearly 2-fold in this experiment and this result has been subsequently reproduced. Addition of EGS or glutaraldehyde was detrimental to specific Pol II and NELF ChIP signals using these and other primers. No enrichment over background was detected using an antibody against HCE with any crosslinking condition, which could be because crosslinking blocked the epitopes recognized by the antibody or because capping enzyme associates transiently with elongation complexes. Although all crosslinked chromatin samples were successfully sonicated to an average size of 300 bp using a Fisher Sonic Dismembrator 550, the amount of recovered DNA was comparatively low from the 1% paraformaldehyde 0.1% glutaraldehyde sample. In a separate experiment, cells crosslinked with 1% glutaraldehyde alone for 1 or 20 minutes either in media or in PBS failed to sonicate—those cell suspensions remained turbid despite extensive sonication and no chromatin was recovered. Overall, these
Figure 6. ChIP-qPCR signal-to-noise at Myc promoter using various crosslinkers
Suspension HeLa cells were crosslinked as indicated. ChIP-qPCR was performed using Pol II, NELF, and HCE antibodies and primer pairs that targeted unoccupied (-3823 bp) and promoter-proximal (+83 bp) regions around Myc. Fold enrichment over background is the ChIP/input signal of the promoter-proximal/unoccupied primer pairs. Error bars represent propagated range (n=2).

results reinforce the importance of controlling crosslinking duration with Tris and demonstrate that longer crosslinking can be beneficial for some ChIP experiments.

Once crosslinking and sonication are optimized to minimize both fragment sizes and epitope destruction, other methods must be used to further improve peak resolution. To this end, I experimented with paired-end reads. If only one end of each fragment is sequenced, the fragment lengths used to determine occupied regions must be estimated either from empirical data (electrophoresis after sonication, profiles of sequencing libraries, etc.) or algorithmically (Jothi et al., 2008; Zhang et al., 2008). By sequencing both ends of each fragment, I reasoned that the signal from fragments shorter than average were more likely to be from specific binding and accurately mapping these fragments would drastically increase resolution. To test this, I performed ChIP-Seq against Pol II, DSIF, and NELF (Liu et al., 2014), which were sequenced using 100 bp paired-end reads, aligned the sequenced fragments using just the first read (single-end read input) or both reads (paired-end read input) (Langmead and Salzberg, 2012), and generated pileup signal using algorithmically estimated or actual fragment length information (Zhang et al., 2008). I then performed metagene analyses, which display the normalized densities of these factors around all RefSeq
Figure 7. Paired-end reads improve ChIP-Seq resolution around promoters
Metagene analysis of Pol II (black), DSIF (blue), and NELF (red) ChIP-Seq occupancies in suspension HeLa cells where sequences were aligned using just the first read (top) or both reads (bottom). Single- or paired-end reads were summed ±10 kb around all RefSeq genes whose transcription start sites (TSS) were not within 1 kb of another TSS; the resulting curves were background subtracted and normalized. The vertical axes represent relative ChIP signal and both plots are the same scale.

genes whose transcription start sites (TSS) are not within 1,000 bp of another TSS (Cheng et al., 2012) (Figure 7). The bidirectional nature of initiation and promoter-proximal pausing (Core et al., 2008; Scruggs et al., 2015) was apparent in profiles generated using paired-end fragment lengths; Pol II, DSIF, and NELF accumulated about 210 bp upstream and 100 bp downstream of the average TSS. These distinct peaks were muddled in the profiles generated using estimated fragment length data. This result clearly shows that paired-end sequencing significantly improves ChIP-Seq resolution.

Elongation control at bidirectional promoters and enhancers
With the increasing resolution of techniques used to detect transcription genome-wide, it was discovered that transcription proceeds bidirectionally from most
promoters (Core et al., 2008; Seila et al., 2008). GRO-Seq (Core et al., 2008), GRO-Cap (Core et al., 2014), and Start-Seq (Scruggs et al., 2015) have all been used to demonstrate that Pol II initiation and transcriptional engagement is divergent in nature around both promoters and enhancers, but higher resolution occupancy data of elongation factors and histone modifications has been lacking. The literature describes H3K4me1 as a marker for enhancers and H3K4me3 as a marker for sites of active transcription (Shlyueva et al., 2014). The role of histone variant H2A.Z, which is also associated with transcriptional activity, is controversial. When H2A.Z substitutes for histone H2A, it is associated with activated genes (Bargaje et al., 2012) and is thought to facilitate Pol II transcription across early nucleosomes (Weber et al., 2014). H2A.Z is also required for enhancer-driven transcriptional induction during differentiation and signaling (Brunelle et al., 2015; Hu et al., 2013). H2A.Z-containing nucleosomes, however, were more stable than canonical nucleosomes (Chen et al., 2013) and have been reported to block transcription by T7 RNA polymerase (Thakar et al., 2010) and HeLa nuclear extracts (Chen et al., 2013). A lingering question in the transcription field is what comes first: transcriptional activity or histone modifications. Also unclear is if chromatin remodeling machinery can differentiate between sense and divergent elongation at promoters and at enhancers. To begin to tackle these questions, I generated high resolution ChIP-Seq datasets against H3K4me1, H3K4me3, and H2A.Z to supplement my Pol II, DSIF, and NELF data in suspension HeLa cells.

First, I performed metagene analyses to gain a broad understanding of the relationship between elongation control machinery (Pol II, DSIF, and NELF) and chromatin (H3K4me1, H3K4me3, and H2A.Z) around RefSeq promoters without neighboring TSS (Figure 8). Pol II, DSIF, and NELF accumulations peaked at both -210 bp and +100 bp relative to the average TSS (Figure 8A). While promoter-proximal pausing has been more accurately mapped to about +50 bp using PRO-Seq (Kwak et al., 2013), the 310 bp distance between sense and antisense paused polymerases in my data is in agreement with a previous report which suggested the median distance between murine divergent TSSs was 176 bp (Scruggs et al., 2015). Because the area under each curve was normalized over a 20 kb window, the relative distributions of factors can be compared. For example, Pol II and DSIF had non-zero signal over the entire sense
Figure 8. Metagene analysis of elongation machinery and histone modifications
Average ChIP-Seq occupancies of Pol II (black), DSIF (blue), and NELF (red) (A-B), or Pol II (black), H3K4me1 (yellow), H3K4me3 (teal), and H2A.Z (pink) (C-D) in suspension HeLa cells. ChIP-Seq reads were summed over the average TSS and the resulting curves were background subtracted and normalized over a 20 kb area. The vertical axes represent relative ChIP signal and (A-B) and (C-D) are the same scale.
direction window (+10 kb), while NELF, which does not associate with productive elongation complexes, fell below Pol II and DSIF at +1 kb and reached background levels at +3 kb (Figure 8B).

Although elongation complexes peaked at -210 bp and +100 bp, the H3K4me3 signal was highly enriched between 100 bp and 2 kb downstream of the average TSS with a peak at about +450 bp (Figure 8C). A small amount of H3K4me3 peaked upstream at -340 bp and this signal tapered to near zero levels by -1 kb. These results suggest that H3K4me3 is strongly associated with productive elongation, which appears to primarily occur in the sense direction. Interestingly, H3K4me1 flanked the average TSS and its signal plateaued at -700 bp and +1.4 kb. While H3K4me1 has been assumed to primarily mark enhancers, this data suggests that active promoters possess a tight accumulation of H3K4me3 surrounded by a relatively lower amount of H3K4me1. The distribution of H2A.Z was slightly biased towards antisense elongation complexes with a major peak at -360 bp and a minor peak at +320 bp. When comparing the volume under the H2A.Z curve around a 4 kb window, 54.1% of H2A.Z was seen upstream (-2 kb to TSS) while 45.9% was seen downstream (TSS to +2 kb). H2A.Z signal also plateaued in the sense direction before H3K4me3, which suggests that H2A.Z substitution might be limited to promoter-proximal histones where H3K4me3 could extend downstream. In contrast with downstream histone modifications, strong H3K4me3 and H2A.Z signals were observed primarily around early antisense nucleosomes. Both H3K4me3 and H2A.Z were relatively depleted beyond 2 kb in both directions (Figure 8D).

As metagene analyses use the average ChIP-Seq signals over many genes, they cannot separate the relative contributions from highly and poorly transcribed promoters. To overcome this limitation, I generated heatmaps for my datasets using the same annotated gene list. I chose to rank-order these TSS by median Pol II occupancy within a 3 kb window because the median would be more likely to reflect the relative Pol II signal in each gene body (an indirect indicator of productive elongation). Although I could have sorted by the maximum Pol II signal over each region (a measure of promoter-proximal pausing), the most paused genes compete with nucleosomes and have less transcription (Gilchrist et al., 2010). H3K4me1, H3K4me3, and H2A.Z occupancies strongly correlated with the presence of transcriptional
Figure 9. Heatmap analysis of elongation machinery and histone modifications

ChIP-Seq occupancies of Pol II (black), DSIF (blue), NELF (red), H3K4me1 (yellow), H3K4me3 (teal), and H2A.Z (pink) in suspension HeLa cells ±1.5 kb around all RefSeq genes whose TSS (dashed line) were not within 1 kb of another TSS. Genes were rank-ordered by the median Pol II signal over each 3 kb area.

machinery and were absent from genes that lacked Pol II (Figure 9). H3K4me3 enrichment was highest at genes with high median Pol II signal and was consistently surrounded by H3K4me1, further emphasizing the relationship between transcription and H3K4 methylation. Although H2A.Z was found in both directions of genes with transcriptional activity, a slight antisense bias was visible at promoters with the highest median Pol II. Unlike H3K4me1 and sense-orientation H3K4me3, H2A.Z occupancy appeared to be spatially limited around TSS in both directions, suggesting that if H2A.Z acts during elongation, it does so early.

To look at ChIP-Seq signals over specific regions of the human genome, I generated compressed pileup files (Kent et al., 2010) and visualized them using the UCSC Genome Browser (Kent et al., 2002). I first examined the highly transcribed DNA-damage-inducible transcript 4 (DDIT4) gene, which is surrounded by enhancer regions. The main peak of promoter-proximal Pol II was about 50 bp downstream of the DDIT4 TSS (Figure 10A). Pol II and DSIF, but not NELF were observed throughout the gene body, indicative of productive elongation, and these factors accumulated 1-2 kb beyond the end of the gene where 3’ processing occurs before termination. Antisense paused
Figure 10. Elongation machinery and chromatin marks around promoters
Pol II, DSIF, NELF, H3K4me1, H3K4me3, and H2A.Z ChIP-Seq occupancy in suspension HeLa cells around the DDIT4 (A) and HSPA5 (B) genes. DDIT4 and RABEPK are transcribed left to right. HSPA5 is transcribed right to left.
polymerases appeared to extend about 1.5 kb upstream of the TSS. H3K4me3 signal was elevated between 350 bp and 2 kb downstream and tapered off before where 3′ processing would begin. As expected from the metagene analyses, a lower amount of H3K4me3 was seen upstream. Additionally, H3K4me1 formed a nest around the H3K4me3-occupied region. H2A.Z was enriched between the TSS and 2 kb upstream and peaked just beyond the region occupied by divergently transcribing Pol II. Although this upstream signal is significantly broader than the averaged metagene distribution, it appears to correlate with antisense and not sense transcription complexes. A number of Pol II, DSIF, and NELF peaks were observed about 10 kb, 13 kb, and 20 kb upstream of DDIT4 and these are suspected enhancers. One of these enhancer polymerases (chromosome 10 position 74,021,000) was associated with high H3K4me3 occupancy and a dip in H3K4me1 signal. Other enhancer peaks had low to undetectable amounts of H3K4me3. Notably, H3K4me1 and H2A.Z were enriched throughout this region.

I next looked the heat shock 70 kDa protein 5 (HSPA5) gene, which is a clear example of a promoter where bidirectional transcription takes place. The main peak of promoter-proximal Pol II was about 110 bp downstream of the HSPA5 TSS and 410 bp away from a distinct antisense paused peak (Figure 10B). Whereas a typical pattern was observed for H3K4me3, H3K4me1 only occupied a region ±5 kb around the HSPA5 gene. Unlike the highly transcribed DDIT4 gene, H2A.Z was present on both sides of the HSPA5 promoter and peaked about 200 bp beyond each Pol II peak. H2A.Z levels also rapidly decreased with distance from HSPA5.

I next selected example regions with known enhancer activity in HeLa cells (Andersson et al., 2014). The NFκB inhibitor alpha (NFKBIA) gene was moderately transcribed and had strong H2A.Z occupancy just upstream of the TSS (Figure 11A). A nest of H3K4me1 was again observed surrounding the H3K4me3-enriched region. A number of enhancer polymerase peaks were observed over the 70 kb region downstream of NFKBIA. As seen in the enhancer region that surrounds DDIT4, regions near transcriptional activity (Pol II, DSIF, and NELF signal) were enriched with both H3K4me1 and H2A.Z. Some peaks also had H3K4me3, but not to the degree seen near promoters of transcribed genes. The basic leucine zipper and W2 domain-containing protein 1 (BZW1), dual specificity protein kinase (CLK1), peptidyl-prolyl cis-trans isomerase-like 3
Figure 11. Elongation machinery and chromatin marks around enhancers
Pol II, DSIF, NELF, H3K4me1, H3K4me3, and H2A.Z ChIP-Seq occupancy in suspension HeLa cells. BZW1 and NIF3L1 are transcribed left to right. NFKBIA, CLK1, and PPIL3 are transcribed right to left. Known enhancer regions are downstream of NFKBIA (A) and upstream of BZW1 (B) (Andersson et al., 2014).
(PPIL3), and \textit{NIF3-like protein 1} (\textit{NIF3L1}) genes, which were transcribed to varying degrees, also had H3K4me3 just downstream of their TSSs, a nest of H3K4me1 surrounding the H3K4me3, and H2A.Z upstream, but not downstream (Figure 11B). Extensive H3K4me1 and H2A.Z, but not H3K4me3 enrichment was seen throughout the enhancer region 10 kb to 100 kb upstream of \textit{BZW1}. Notably, the 30 kb region between \textit{BZW1} and \textit{CLK1}, a region where Pol II initiation is not observed, was depleted of these histone modifications.

These results support a model where H3K4me1 is deposited anywhere near sites of Pol II transcription. If these polymerases routinely enter productive elongation, this H3K4 monomethylation is “upgraded” to trimethylation and maintenance of this mark requires continued productive elongation and PAF1 complex recruitment. Because monomethylation is an intermediate towards trimethylation, a nest of H3K4me1 will be formed around well-maintained regions of H3K4me3. When the transition to productive elongation is infrequent—such as at enhancers, divergent promoters, and infrequently transcribed genes—initiation and pausing could be sufficient to generate H3K4me1 but not enough to maintain high H3K4me3. While H3K4 methylation appears to be incidental to transcription, the timing of H2A.Z substitution remains unclear. Although H2A.Z is present only in regions of ongoing transcription and is tightly positioned around promoters—which would suggest its deposition is in association with transcription—it broadly occupies enhancer regions. H2A.Z also seems to anticorrelate with productive elongation downstream of some, but not all promoters. Overall, these data suggest that sense, antisense, and enhancer elongation and H3K4 methylation machineries are fundamentally similar and likely depend on regular productive elongation, which is influenced by promoter strength and input from transcription factors and other regulatory signals. The role of H2A.Z, however, remains unknown.

\textit{Transcription in a primary T cell model of HIV latency}

In HIV-infected patients, the bulk of latent virus persists in reservoirs of infected central memory T cells (Chomont et al., 2009). Although other reservoirs exist in other cell types, only memory T cells have been demonstrated thus far to persist in an infected state for years during ongoing antiretroviral treatment (Churchill et al., 2016). HIV latency has been shown to be influenced by viral mutation and integration
site selection (Ho et al., 2013), DNA methylation (Blazkova et al., 2012; Kauder et al., 2009), repressive chromatin (Pearson et al., 2008; Tyagi et al., 2010), and transcriptional interference from nearby genes (Greger et al., 1998; Han et al., 2008; Lenasi et al., 2008).

A number of models exist to study transcription during HIV latency (Bosque and Planelles, 2011; Tyagi and Romerio, 2011). Transformed T cell lines that were cloned after HIV infection are readily grown in sufficient quantities for sequencing experiments; in fact, the only two published ChIP-Seq experiments over the HIV genome use these models (Jadlowsky et al., 2014; Park et al., 2014). Because transformed cells actively divide, however, they cannot mimic the low metabolism and lack of proliferation seen in quiescent memory T cells (Hua and Thompson, 2001). Another disadvantage is each individual cell line has its own individual site of HIV integration. In latently infected patient cells, differences in these sites of integration resulted in differences in viral reactivation (Ho et al., 2013). Infected patient cells would accurately portray this range of transcriptional effects, but only about one copy of HIV is present per million resting CD4-positive T cells (Ho et al., 2013); this number is too low for ChIP-Seq to be feasible.

Several primary cell culture models have been developed where primary CD4-positive T cells are isolated and infected with HIV (Bosque and Planelles, 2009; Marini et al., 2008; Sahu et al., 2006; Saleh et al., 2007; Tyagi et al., 2010; Yang et al., 2009). One such model isolates naïve CD4-positive T cells to near purity, and then activates and differentiates them to resemble central memory cells (Bosque and Planelles, 2009, 2011; Martins et al., 2016). The majority of these cells can be infected and subsequently suppressed to only 2% viral activity. A collaboration was started with Alberto Bosque and Vicente Planelles, who generated these cells and provided otherwise unobtainable samples. With this primary T cell model of HIV latency and reactivation, I generated Pol II ChIP-Seq datasets in an attempt to study transcription over the HIV promoter.

To confirm that this model accurately mimicked memory T cell quiescence (the maintained state of non-proliferation and low metabolism) and reactivation to stimuli, I scanned the genome for genes whose Pol II occupancy changed after cellular stimulation with beads coated with αCD3 and αCD28 antibodies (Figure 12). The Krüppel-like factor 2 (KLF2) gene is highly expressed in quiescent T cells and is rapidly
downregulated in response to T cell activation (Preston et al., 2013). As expected, Pol II occupancy over KLF2 was high only in quiescent cells and was nearly undetectable after activation (Figure 12A). MIR155HG, which is highly expressed after T cell activation (Elton et al., 2013), also showed expected patterns of Pol II occupancy (Figure 12B). Importantly, the presence or absence of HIV infection had little effect on Pol II occupancy genome-wide. These results indicate that these cells properly model the transcriptional behaviors of central memory T cells.

To visualize ChIP-Seq occupancies over the HIV genome, I developed and implemented a paired-end alignment strategy that reduced ambiguous assignment of fragments to the duplicated HIV long terminal repeats. I assembled a pNL4-3 reference genome that contained appropriately transposed LTRs, which occurs normally during reverse transcription. Because paired-end sequencing was performed, fragments where portions of one end mapped in a unique region (within the coding region of HIV or within the human genome) could be unambiguously assigned (Figure 13, black signal). As sequenced fragments had an average length of about 300 bp, some were not able to be uniquely assigned and these are duplicated over each LTR (Figure 13, red or blue signal). In quiescent cells, Pol II had a generally uniform occupancy over the HIV genome with slight enrichment over the LTR regions. This pattern could be the result of non-specific background signal or from relatively uniform transcription originating from either the 5′ LTR (due to incomplete silencing), or from nearby cellular promoters. Activation increased Pol II occupancy across the virus and over both LTRs. Notably, more sequencing reads could be unambiguously assigned to the 5′ LTR after activation. Curiously, I detected a small number of reads that aligned precisely 2 bp upstream of the 3′ LTR. Because these reads originated only from one strand (only 5′–AC was observed; the reverse compliment 5′–TG was not detected), I concluded that they were from unintegrated linear viral DNA. Raltegravir, an integrase inhibitor, was used to suppress HIV after infection in this experiment and is the likely cause of these unintegrated products. By determining the genome size-adjusted ratio of HIV-mapped to human-mapped reads in input samples, I also found that these cells were infected with 3.6–4.2 copies each, a copy number which could potentially prevent latency. This experiment demonstrated that ChIP-Seq could be used to examine the occupancy of Pol
Figure 12. Pol II occupancy at genes differentially expressed in T cells ± αCD3/αCD28

Pol II ChIP-Seq occupancy around KLF2 (A) or MIR155HG (B) in uninfected or HIV-infected primary central-memory-like T cells that were quiescent or activated for 24 h with beads coated with αCD3 and αCD28 antibodies. All tracks were normalized using the average number of hg19-mapped sequence reads.
Figure 13. ChIP-Seq occupancy of Pol II over the HIV genome ± αCD3/αCD28
Pol II ChIP-Seq occupancy over the HIV genome in primary central-memory-like T cells that were quiescent or activated for 24 h with beads coated with αCD3 and αCD28 antibodies. Reads which could not be definitively assigned to either LTR are duplicated in both locations (5′: red; 3′: blue).

II and other factors over the HIV genome. Based on these results, changes were made to the infection model to reduce unintegrated products and lower HIV copy numbers.

Targeted reactivation of HIV by STAT5 after HODHBt treatment
In a screen for chemicals that reactivated latent HIV in memory T cells, Alberto Bosque identified 1-hydroxybenzotriazol or HOBt. By testing structurally similar compounds, he discovered that 3-hydroxy-1,2,3-benzotriazin-4(3H)-one (HODHBt) could induce latent HIV without reactivating otherwise resting T cells in both infected patient cells and in his primary T cell model of HIV infection. HODHBt was found to stabilize the active, phosphorylated form of STAT5, a transcription factor which regulates the IL-2 response in T cells (Lin and Leonard, 2000). Additionally, inhibition of STAT5 but not STAT3 abrogated the reactivation effects of HODHBt. To determine if STAT5 had a direct role in HIV reactivation, I performed ChIP-Seq in latently-infected primary T cells that were untreated or cultured with IL-2 ± 100 μM HODHBt for 3 hours against Pol II and STAT5A, the primary STAT5 isoform in these cells.

In all three conditions, Pol II had a generally uniform occupancy over the HIV genome (Figure 14). However, a distinct peak of STAT5A appeared at position 329 after HODHBt treatment and this peak was positioned between two STAT5 dimer binding sequences that are an appropriate distance apart to enable STAT5 tetramerization (HIV318-350: TTCAAGAA-N15-TTGCTACAA) (consensus: TTCYNRGAA-N2-15-TTCYNRGAA)
Figure 14. ChIP-Seq occupancies of Pol II and STAT5A over the HIV genome
Pol II and STAT5A ChIP-Seq occupancy over the HIV genome in primary central-memory-like T cells that were untreated or treated with IL-2 ± 100 μM HODHBt for 3 h. All tracks were normalized using the average number of hg19-mapped sequence reads. Reads which could not be definitively assigned to either LTR are duplicated in both locations (5’: red; 3’: blue). The sequence of HIV NL4-3 position 318-350 is indicated and compared with the STAT5 tetramer consensus sequence (Lin et al., 2012b). This result indicates that STAT5A is specifically recruited to the HIV LTR promoter after HODHBt treatment and supports a model where STAT5 can directly reactivate HIV. In an improvement from the prior experiment, these cells had fewer unintegrated viral genomes and hosted fewer than 2 copies per cell.

To determine if HODHBt reactivates T cells and to identify its gene targets, I next compared the effects of HODHBt and αCD3/αCD28 stimulation by stranded RNA-
Seq. Cultured central-memory-like T cells from three individual donors were divided equally and treated for 24 hours with IL-2 alone (control), with IL-2 and 100 μM HODHBt, or with αCD3/αCD28-coated beads. After Trizol extraction, polyadenylated (polyA) RNAs were enriched and prepared for paired-end sequencing using a kit that randomly ligates one of 96 possible sequencing adapters to each end—this enabled separation of abundant RNA fragments with otherwise identical sequences from duplicate PCR amplification products with 9,216-to-1 confidence. After alignment using TopHat (Kim et al., 2013) and PCR duplicate removal using a custom script, I followed a pipeline developed by Alec McIntosh and Miles Pufall. First, I performed feature counting using Subread, which quantifies reads in a manner independent of transcript length (Liao et al., 2014). These feature counts were then imported into R and normalized using both EDASeq (Risso et al., 2011), which reduces variability from different sequencing depths between replicates, and RUVSeq (Risso et al., 2014), which further reduces variability from batch effects without the use of control genes. Differential expression analysis was then performed using DESeq2 (Love et al., 2014) and settings which prevented additional normalization. False discovery rate-adjusted p-values (q-values) and a conservative cutoff of 0.01 were used to identify genes whose expression was significantly altered after either treatment. Similar results were also obtained using a well-established Cufflinks pipeline (Trapnell et al., 2012).

First, normalized gene expression values were averaged between the three donors. When these averages were compared between HODHBt and control cells, only 394 of 26,364 genes were significantly altered (Figure 15, red dots). T cell reactivation by αCD3/αCD28 treatment, however, significantly changed the expression of nearly a third of all genes (Figure 15, blue dots). The changes seen with HODHBt were also markedly lower in magnitude than with αCD3/αCD28 as indicated by the relative distances of colored dots from the $x=y$ diagonal (which represents no change). These results suggest that HODHBt treatment causes relatively minimal changes in cellular gene expression. To better visualize the expression differences between these treatments, I plotted the log$_2$ fold changes of genes significantly altered by one or both treatments (Figure 16). Where the most changed genes after HODHBt treatment were GBP1P1 (2.95-fold increase) and RGCC (1.80-fold decrease), changes in gene expression
Figure 15. Differential expression in T cells after HODHBt or αCD3/αCD28
Differential expression analysis of polyA-selected RNA-Seq from 3 donors in matched primary central-memory-like T cells that were treated for 24 h with IL-2 alone (control), IL-2 and 100 μM HODHBt, or αCD3/αCD28-coated beads. False discovery rate-adjusted p-values (q-values) were used to identify genes whose expression was significantly altered after HODHBt (red) or αCD3/αCD28 (blue) treatment. Non-significant genes (q > 0.01) are shown at 5% opacity (black).

Figure 16. Differential expression of genes significantly changed by HODHBt and/or αCD3/αCD28
Log₂ fold changes (treatment vs. control) of genes whose expression was significantly altered by HODHBt only (red), αCD3/αCD28 only (blue), or both treatments (purple) were compared.
after αCD3/αCD28 treatment were much higher in magnitude, the largest of which were *IL2* (4,140-fold increase) and *KLF2* (304-fold decrease). About 82% of genes with altered expression after HODHBt were also significantly changed by αCD3/αCD28 (Figure 16, purple dots). While most of these shared genes that were downregulated by HODHBt also went down after αCD3/αCD28 treatment (156/171), genes upregulated by HODHBt were generally not co-regulated by αCD3/αCD28 (101/223). Because differential expression analysis relies on a number of software manipulations of feature counts, I verified these findings by visualizing the normalized and combined RNA-Seq pileups over the most changed genes after only HODHBt (Figure 17), only αCD3/αCD28 (Figure 18), or both treatments (Figure 19). Not only were similar results seen in these relatively unmanipulated pileups, the high quality of the data was demonstrated by the nearly undetectable amount of background signal. Pathway analysis of these gene sets further indicated that HODHBt uniquely activated STAT-family transcription factors and αCD3/αCD28 uniquely induced markers of T cell activation (Roider et al., 2009). Overall, these results demonstrate that HODHBt can induce latent HIV in T cells through STAT5 without triggering T cell activation. Additional results obtained by Alberto Bosque also confirmed that HODHBt works on latently infected cells isolated from HIV patients and had minimal effects on CD69 and CD25 surface protein levels, which are normally elevated after T cell activation.

*Possible regulation of HEXIM1 by transcriptional interference*

Expression of *HEXIM1*, the gene which encodes the component of the 7SK snRNP responsible for inhibiting sequestered P-TEFb, is induced after chemical insults that cause release of P-TEFb from the 7SK snRNP (Liu et al., 2014). This increased expression of *HEXIM1* is thought to help restore the balance of free and sequestered P-TEFb, thus maintaining transcriptional homeostasis (He et al., 2006). By performing ChIP-Seq against Pol II, DSIF, and NELF in suspension HeLa cells, I demonstrated that promoter-proximal pausing occurs about 40 bp downstream of a previously unannotated *HEXIM1* promoter, which is the origin of *HEXIM1* transcription after induction (Liu et al., 2014). In that study, *HEXIM1* but not *HEXIM2* (which is about 9 kb downstream) was inducible by the histone deacetylase inhibitor SAHA (Liu et al., 2014). An unanswered question remained from my ChIP-Seq data over the region surrounding *HEXIM1*—why was Pol II
Figure 17. RNA-Seq signal over genes significantly changed only by HODHBt
Aligned reads from polyA-selected RNA-Seq from 3 donors in matched primary central-memory-like T cells that were treated for 24 h with IL-2 alone (control), IL-2 and 100 μM HODHBt, or αCD3/αCD28-coated beads. Reads from individual donors were first normalized using the average number of hg19-mapped sequence reads and then combined. (A) GBP1P1 increased 2.95-fold after HODHBt. (B) RGCC decreased 1.80-fold after HODHBt.
Figure 18. RNA-Seq signal over genes significantly changed only by αCD3/αCD28
Combined, normalized reads from polyA-selected RNA-Seq in primary T cells treated for 24 h with IL-2 alone (control), IL-2 and 100 μM HODHBt, or αCD3/αCD28. (A) IL2 increased 4,140-fold after αCD3/αCD28. (B) KLF2 decreased 304-fold after αCD3/αCD28.
Figure 19. RNA-Seq signal over genes significantly changed by both treatments

Combined, normalized reads from polyA-selected RNA-Seq in primary T cells treated for 24 h with IL-2 alone (control), IL-2 and 100 μM HODHBt, or αCD3/αCD28. (A) SOCS3 increased 3.07-fold after HODHBt and 2.42-fold after αCD3/αCD28. (B) GAL3ST4 decreased 2.79-fold after HODHBt and 22.2-fold after αCD3/αCD28; GPC2 decreased 2.55-fold after HODHBt and 6.86-fold after αCD3/αCD28.
and DSIF occupancy observed throughout the gap between HEXIM1 and HEXIM2? With the generation of histone ChIP-Seq and stranded RNA-Seq tracks, I revisited the HEXIM1 promoter to determine if runaway transcription (elongation beyond the site of termination) or transcriptional interference (convergent transcription from a distant downstream promoter) were occurring.

The previously reported peak of Pol II, DSIF, and NELF 40 bp downstream of the unannotated HEXIM1 TSS was flanked by H3K4me1 and, as expected from productive elongation from left to right, H3K4me3 occupied a region over the next 2-3 kb downstream (Figure 20). H2A.Z was highly enriched upstream of the HEXIM1 TSS. Two peaks of paused Pol II were observed about 425 bp apart within the HEXIM2 gene. The left peak is about 425 bp upstream of an alternative HEXIM2 TSS (not shown); the right peak is about 50 bp downstream. Surprisingly, H3K4me3 was highly enriched on the left side of these polymerase peaks, which suggests that productive elongation is biased in the antisense direction from HEXIM2. I next looked at strand-specific RNA-Seq pileups of the polyA-enriched RNAs from primary T cells treated with IL-2 alone, IL-2 with 100 μM HODHBt, or with αCD3/αCD28 antibody beads. Strong accumulations of sense-strand mRNA were observed over the HEXIM1 gene. Antisense-strand RNA-Seq signal was also seen throughout the region between HEXIM1 and the upstream HEXIM2 peak of Pol II. Poor alignment through repetitive sequence elements accounts for a gap in coverage observed just downstream of HEXIM1 (Figure 20, RepeatMasker). Because this region contains stretches of adenine and thymine base repeats, it is unclear if the polyA-enriched antisense long noncoding RNA (lncRNA) originating from HEXIM2 was truly polyadenylated or was selected because of these repetitions. Importantly, αCD3/αCD28 treatment induced HEXIM1 gene expression about 2-fold and this increase in sense RNA signal was accompanied by a 6-fold decrease in antisense transcription originating from HEXIM2. These results demonstrate that antisense transcription originating from HEXIM2 spans into HEXIM1, and the abundance of antisense HEXIM2 and sense HEXIM1 transcripts is inversely correlated. The degree of overlap between these complementary transcripts could have significant regulatory consequences. Although sense HEXIM2 transcription went up about 2-fold after both HODHBt and
Figure 20. Possible regulation of HEXIM1 gene expression by antisense transcription
Pol II, DSIF, NELF, H3K4me1, H3K4me3, and H2A.Z ChIP-Seq occupancy in suspension HeLa cells and combined, normalized reads from polyA-selected RNA-Seq in primary T cells treated for 24 h with IL-2 alone (control), IL-2 and 100 μM HODHBt, or αCD3/αCD28. RepeatMasker annotates genomic regions of interspersed repeats and low complexity DNA.
\(\alpha CD3/\alpha CD28\), only \(\alpha CD3/\alpha CD28\) downregulated antisense \(HEXIM2\) transcription, which suggests that sense and antisense \(HEXIM2\) transcription may separately regulated.

*Sequence specificity poorly defines the genomic occupancy of Myc*

Myc has been shown to generally occupy promoter and enhancer regions genome-wide (Consortium, 2012; Lin et al., 2012a) and is present near the transcription start sites of a third of actively transcribed genes in embryonic stem cells (Rahl et al., 2010). Two recent studies that used different cell types also demonstrated that Myc universally induced the existing gene expression programs (Lin et al., 2012a; Nie et al., 2012). These studies, however, attributed the observed universal effects of Myc to either an enrichment of accessible high-affinity E-box variants at core promoters and enhancers (Lin et al., 2012a) or the ubiquitous nature of degenerate E-box sequences (Nie et al., 2012). If Myc activity requires close proximity to its targets, sequence specificity is unlikely to exclusively define where Myc binds. Supporting this hypothesis, ChIP-Seq experiments performed in numerous cell types (Consortium, 2012) show that Myc and Max bind DNA near Pol II transcription machinery and regularly occupancy regions that lack E-box sequences (Guo et al., 2014a). To help resolve this conflict, I performed analyses which incorporated both the in vitro binding preferences of Myc-Max and the actual occupancy of Myc across the genome.

To determine the binding preferences of Myc-Max across all possible 8 bp sequences (or 8-mers), I first analyzed data generated by protein-binding microarray (PBM) assays performed by Joshua Schipper and Raluca Gordân (Guo et al., 2014a) that used “all 10-mer” universal array designs (Berger and Bulyk, 2009; Berger et al., 2006). After normalization, I found that the relative Myc-Max occupancy over each of the 32,896 8-mers exhibited a 56-fold range (from 0.018 to 1) (Figure 21, inset). This small range suggests that Myc-Max binds somewhat indiscriminately with DNA, a hypothesis that was supported by electrophoretic mobility shift assays (Guo et al., 2014a). Most of the sequences containing the canonical E-box sequence CACGTG had high occupancy, but flanking bases had a significant influence on Myc-Max binding (Figure 21). Several E-box variants, the top 12 of which are shown in Figure 21, were also significantly affected by flanking bases. Even after establishing this expanded Myc-Max “binding
Fluorescent signals generated by Myc-Max in vitro binding to an array containing all possible 8-mers were normalized. Twelve core 6-mer sequences with the highest in vitro occupancy are shown. The relative affinity of all 8-mers for each 6-mer is shown (10 points if the 6-mer is a palindrome or 16 if it is not). The inset shows the sorted in vitro binding signal for all possible 8-mers.

spectrum” of non-canonical E-boxes, the genomic locations of Myc and these sequences did not appear to be strongly correlated (Guo et al., 2014a).

I next performed analyses to calculate the correlation between the in vitro sequence preferences of Myc-Max and the actual genomic occupancy of Myc in HeLa cells. A 100 bp interval around the top 30,000 Myc ChIP-Seq peaks, as determined by Tiandao Li using ChIP-Seq Peak (Cheng et al., 2012), was scanned to find the 8-mer with the highest possible in vitro binding affinity. Peaks were then scored using the normalized in vitro affinities of the best 8-mer, rank-ordered from highest to lowest values, and plotted (Figure 22, blue dots). Seventy-four percent of the top 30,000 Myc peaks were associated with low affinity 8-mers with normalized in vitro occupancies below 0.2. To determine if the distribution of 8-mers around sites of Myc occupancy was different from what would occur by chance, I performed the same analysis on 30,000 100 bp regions randomly chosen from DNase I-sensitive regions (Thurman et al., 2012) (Figure 22, black dots). DNase I-sensitive regions were chosen as a control for this
Figure 22. Sites of Myc occupancy in cells are slightly enriched for quality E-boxes
The top 30,000 sites occupied by Myc (blue) were rank-ordered and scored by the in vitro binding affinity of the best 8-mer in a 100 bp window (y-axis). This was repeated at 30,000 random locations of DNase I-sensitivity (black) and the results were directly compared by ROC analysis (inset).

This slight but significant correlation was also visualized with a receiver operating characteristic (ROC) curve (Figure 22, inset). ROC curves are commonly used in genomic analyses to assess whether a specific quantitative feature (here, normalized in vitro Myc occupancy scores) can distinguish between two classes of sequences (here, ChIP-Seq peaks versus random accessible regions). Although the area under the ROC curve was better than expected by chance (0.637 vs. 0.5), this data suggests that Myc-Max in vitro occupancy scores are a poor predictor of whether or not an accessible genomic region will be bound by Myc in cells. For example, at a false positive rate (1-specificity) of 10%, the true positive rate (sensitivity) of identifying a genomic region occupied by Myc from an in vitro occupancy score is only about 25%. This analysis demonstrates that the vast majority of sites occupied by Myc are associated with low affinity E-box sequences.
To determine if the intrinsic binding specificity of Myc-Max influences the degree of its genomic occupancy in cells, I rank-ordered the same 30,000 Myc peaks by their ChIP-Seq signal and directly compared the normalized occupancy at each peak with the normalized score of the best 8-mer ±50 bp. A 30-fold range in ChIP-Seq occupancies was observed across the top 30,000 Myc peaks (Figure 23A, blue line). Using the same x-axis, a second plot was generated to display the relative affinity of the best 8-mer associated with each of these Myc peaks (Figure 23A, black dots). A slight preference for high affinity 8-mers was observed over the top 5,000 Myc peaks, but the overwhelming conclusion is that 8-mers with a wide range of in vitro occupancies are found around Myc peaks irrespective of the level of in vivo occupancy (Figure 23A). Although a statistically significant correlation was observed between Myc ChIP-Seq occupancy and in vitro 8-mer binding strength, this relationship was weak (Spearman correlation coefficient: \( \rho = 0.22, p\text{-value} < 2.2 \times 10^{-16} \)). Had the cellular occupancy correlated well with the affinity for the underlying DNA sequences, a cloud of black dots would be clustered around the blue curve in Figure 23A and the Spearman correlation coefficient would have been closer to 1. A plot of the same data after ChIP-Seq peaks were grouped into log-scaled bins provided a more detailed view of the high occupancy sites in cells that might be expected to correlate better with intrinsic DNA affinities; however, a huge range of in vitro occupancy scores was observed even for the highest occupancy sites (Figure 23B). Overall, these results suggest that Myc occupancy is driven only to a small extent by its intrinsic sequence preference and additional mechanisms must be required to recruit Myc to its genomic binding locations in cells.

**Methods to detect and isolate Pol II nascent transcripts**

Although ChIP-Seq can be used to measure the location of Pol II in cells, the most direct method of measuring transcriptionally engaged polymerases is by nuclear run-on. In this method, nuclei are gently isolated and nascent transcripts are extended in the presence of sarkosyl to prevent new initiation from occurring. Nascent transcripts can be labeled either through the use of radioactive nucleotides or with nucleotide analogues that enable subsequent purification and sequencing. In GRO-Seq (Core et al., 2008), nuclear run-ons are performed in the presence of 5-bromouridine 5'-
Figure 23. Myc ChIP-Seq occupancy levels weakly correlate with E-box quality
(A) The top 30,000 sites occupied by Myc were rank-ordered by ChIP-Seq signal and scored by either normalized ChIP-Seq signal (blue line) or the in vitro binding affinity of the best 8-mer in a 100 bp window (black dots). (B) The data in (A) are presented using a default R boxplot (box: 1st to 3rd quartile, line: median, whiskers: 1.5 × interquartile range beyond the box, outliers are stacked) with ChIP-Seq signal in blue and in vitro 8-mers in grey.
triphosphate (BrUTP), which is inexpensive but fails to stop further elongation. PRO-Seq (Kwak et al., 2013) instead uses biotin-11-NTPs, which while expensive, only incorporate once per nascent transcript enabling single nucleotide precision. In anticipation of using these techniques in the future, I performed trial nuclear run-ons to compare the properties of both BrUTP and biotin-11-NTPs (Figure 24). I also isolated these labeled transcripts with either M-280 Streptavidin Dynabeads (Figure 24, lanes 1-6) or anti-BrdU agarose beads (Figure 24, lanes 7-10) to evaluate their respective binding efficiencies.

A control nuclear run-on was required to evaluate if run-ons performed with biotin-11-NTPs stopped after one incorporation, so only $\alpha\text{-}^{32}\text{P}-\text{CTP}$ was added to the first elongation reaction. In the absence of ATP, UTP, and GTP, engaged polymerases should only move forward if the next nucleotide to be incorporated is CTP—the odds of this are slightly greater than 1:4 due to the GC-rich nature of promoters and early exons (Louie et al., 2003). The vast majority of polymerases which elongate in this condition should only move forward one nucleotide. A population of transcripts was observed between 30-50 nt, typical for promoter-proximal pausing (Figure 24, lane 2). Importantly, no transcripts bound non-specifically to the streptavidin beads (Figure 24, lane 1), which suggests that binding and washing conditions were sufficiently stringent. An additional RNA population was seen around 20 nt. I will later show in Figure 26 that these RNAs, once isolated, were unresponsive to calf intestinal alkaline phosphatase or high amounts of HCE, suggesting that they were the result of 5′ degradation to the edge of the polymerase. This 20 nt population was also reduced in future experiments when I isolated nuclei more rapidly and in the presence of EDTA (Chapter 4). I termed run-ons performed with only $\alpha\text{-}^{32}\text{P}-\text{CTP}$ as “nuclear walk-ons” because of their short polymerase elongation distance. Nuclear walk-ons are now routinely used in the Price Lab because of their ability to obtain a metagene-like view of polymerase positions without the expense or delay of sequencing.

Next, I performed nuclear run-ons with limiting $\alpha\text{-}^{32}\text{P}-\text{CTP}$ and 10 $\mu$M each of biotin-11-ATP, -UTP, and -GTP, a concentration 37.5-fold less than used in prior studies (Kwak et al., 2013; Schaaf et al., 2013). Trizol-isolated nascent transcripts were precipitated in ethanol containing 0.5 M ammonium acetate once (Figure 24, lanes 3-4).
Figure 24. Nuclear run-ons with α-³²P-CTP only, biotin-11-NTPs, or BrUTP
Nuclear run-ons performed with nuclei isolated from suspension HeLa cells. After elongation in the presence of NTPs as indicated, Trizol-isolated nascent transcripts were pulled down with either M-280 Streptavidin Dynabeads (lanes 1-6) or anti-BrdU agarose beads (lanes 7-10). B: bound transcripts. S: pooled unbound and wash supernatants. 9% Urea-PAGE.
or twice (Figure 24, lanes 5-6) to determine if further reduction of excess unincorporated biotin-11-NTPs would improve affinity purification. In both reactions, about a third of all labeled nascent transcripts were bound to the streptavidin beads. All labeled transcripts also shifted up uniformly, suggesting that only one biotin-11-NTP was incorporated. If multiple biotin-11-NTPs were incorporated, I would expect the discreet, biotin-shifted band at ~50 nt to be smeared. The lack of an unshifted band (~48 nt) also suggests that polymerases efficiently incorporated biotin-11-NTPs at 10 µM. While the efficiency of binding was unaltered by an additional precipitation, it could potentially be improved by increasing the amount of beads used in future experiments. This result indicates that biotin-11-NTPs can be used at lower concentrations for nuclear run-ons, reducing the expense to about $8 per reaction.

To evaluate how much elongation was required to efficiently isolate BrUTP nascent transcripts, I performed nuclear run-ons with all four nucleotides present at very low concentrations (Figure 24, lanes 7-8), or at higher concentrations previously used for GRO-Seq (Figure 24, lanes 9-10) (Core et al., 2008). As expected, increasing the NTP concentration increased both the length of all nascent transcripts (due to less CTP starvation) and the fraction of transcripts bound to the anti-BrdU beads (due to more BrUTPs per transcript). The fraction of BrUTP transcripts bound, however, was much lower than the fraction bound after biotin-11-NTP transcript purification. Additionally, neither BrUTP condition provided the resolution attained with biotin-11-NTPs or α-³²P-CTP only.

To better understand the incorporation properties of biotin-11-NTPs, I introduced them to in vitro transcription reactions that reconstitutes initiation, pausing, and coupled RNA processing (Adamson et al., 2003). In this system, HeLa nuclear extract and a CMV-promoter-driven template are preincubated for 30 minutes to allow formation of preinitiation complexes. Initiation is then accomplished in a 30 second pulse with limiting α-³²P-CTP. Factors associated with these early elongation complexes are then removed by high salt wash. At this point, I either isolated radiolabeled transcripts (Figure 25, lanes 1-2) or incubated washed complexes for 3 minutes with 10 µM biotin-11-NTPs (Figure 25, lanes 3-6). In the control reaction, a typical distribution of transcripts was generated by a 30 second limiting-CTP pulse and
30 s pulse, then HSW

<table>
<thead>
<tr>
<th>3 min incubation</th>
<th>1 min chase</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 μM Bio-11-A/U/G</td>
<td>500 μM A/U/G/C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>S</th>
<th>B</th>
<th>S</th>
<th>B</th>
<th>S</th>
</tr>
</thead>
</table>

**Figure 25. Incorporation of biotin-11-NTPs by in vitro elongation complexes**

HeLa nuclear extract and template DNA were preincubated for 30 min and pulsed for 30 s with limiting α-³²P-CTP. Elongation complexes (ECs) were then high salt washed (HSW). In lanes 1-2, transcripts from ECs were isolated immediately. In lanes 3-6, ECs were incubated for 3 min with 10 μM biotin-11-NTPs. In lanes 5-6, ECs were also chased for 1 min with 500 μM NTPs. All RNAs were affinity purified with M-280 Streptavidin Dynabeads. B: bound. S: pooled unbound and wash supernatants. 9% Urea-PAGE.
almost no non-specific binding of these RNAs was observed (Figure 25, lane 1). While incorporation of biotin-11-NTPs was incomplete in vitro, separation of biotinylated RNAs from undecorated transcripts was very efficient (Figure 25, lanes 3-4). Inefficiencies in biotin-11-NTP incorporation could be due to the absence of sarkosyl, which is present during nuclear run-ons but not in vitro transcription reactions. Biotin-11-NTP incorporation resulted in a shift equivalent roughly to a 3 nt addition in agreement with the previous nuclear run-on. To determine if biotin-11-NTP blocks further elongation as previously reported (Kwak et al., 2013), elongation complexes preincubated with 10 μM biotin-11-NTPs for 3 minutes were chased for 1 minute with 500 μM of ATP, UTP, GTP, and CTP (Figure 25, lanes 5-6). Unexpectedly, long transcripts were observed in both bound and supernatant lanes, indicating that biotin-11-NTPs do not stop incorporation of NTPs. Because this result could also be obtained if biotin-11-NTPs were incorporated after further elongation had occurred, a wash step could be added before the chase in a future experiment. Fortunately, carefully isolated nuclei should lack endogenous NTPs and unwanted elongation was not observed in Figure 24.

In two nuclear walk-on experiments, I observed a population of 20 nt RNAs associated with elongation complexes (Figures 24 and 39). Because the 5’ end of nascent transcripts 20 nt in length would be positioned at the threshold of the polymerase RNA exit pore, it was possible that their accumulation was evidence for a regulated “capping pause” before the typical 30-50 nt pause associated with DSIF and NELF. The alternative was that they were the result of 5’-to-3’ RNA degradation to the edge of the polymerase. RNAs degraded in this manner would lack 5’ triphosphate ends. To address this question, I treated transcripts isolated from in vitro transcription reactions (Figure 26, lanes 1-6) and nuclear walk-ons (Figure 26, lanes 7-12) to completion over 30 minutes with 25 units of calf intestinal alkaline phosphatase (CIP) or 1 μg of human capping enzyme (HCE). Because cap addition causes a small mobility shift (equivalent to about 1.5 nt) which is challenging to detect in the absence of discreet bands, I separated capped and uncapped RNAs after CIP or HCE treatment by further treating them with human cap methyltransferase (HCM) and SAM to mature existing caps, and then immunoprecipitating them with anti-2,2,7-trimethylguanosine beads (Moteki and Price, 2002; Nilson et al., 2015).
After a 30 second pulse, a typical pattern of transcripts was generated (elongation complexes stop after 16, 21, 25, 31, 49, 54, and 56 bp using this template and limiting CTP). Capped and uncapped forms of all transcripts greater than 20 nt were observed; because transcripts shorter than 20 nt do not extend beyond the body of the polymerase, they cannot be capped co-transcriptionally (Coppola et al., 1983). Transfer RNAs (tRNA) also incorporate CTP in the presence of extract independently of transcription and should be ignored. When CIP was added to these transcripts, uncapped transcript mobilities decreased; this mobility shift corresponds with the loss of negative charges from 5′ triphosphate cleavage. Capped transcripts were unaffected because the inverted guanosine cap protects these phosphates. When HCE and GTP were added, previously uncapped transcripts were capped as evidenced by a small decrease in mobility and nearly complete binding to the anti-2,2,7-trimethylguanosine beads. Because G16, a transcript too short to be capped, was affected by CIP and HCE treatment, I could determine if the 20 nt nascent transcript population in my nuclear walk-on experiments possessed 5′ triphosphate ends.

A fraction of nuclear walk-on transcripts associated with paused polymerases (around 30-50 nt) were capped. The 20 nt transcripts, however, were not detected in the bound fraction, indicating that they were generally uncapped. Because uneven amounts of material were loaded on the gel (as evidenced by ethidium bromide staining of small nuclear RNAs), I could not quantify the ratio of capped-to-uncapped transcripts. After CIP treatment, the 20 nt population did not appear to shift. While a slight change in mobility may have occurred in uncapped 30-50 nt walk-on RNAs, a strong comparison could not be made due to difficulties in detecting this shift in the 31 nt and 49-56 nt in vitro transcripts. A clear result was obtained using HCE; the 20 nt population of isolated nuclear walk-on transcripts were not capped by a 30 minute incubation with 1 μg of capping enzyme. The lack of capped 20 nt signal cannot be explained by poor anti-2,2,7-trimethylguanosine binding efficiency as recovery of capped Pol II-transcribed snRNAs was complete. Because these ~20 nt nascent transcripts could not be capped and appears unresponsive to CIP, I conclude that they were generated by 5′ degradation to the edge of the polymerase.
Figure 26. Responsiveness of transcripts to phosphatase or capping enzyme

(Lanes 1-6) Transcripts generated by a 30 s pulse with limiting α-³²P-CTP. (Lanes 7-12) Transcripts generated by nuclear walk-ons. Isolated transcripts were incubated with buffer only (-), 25 U CIP, or 1 μg of HCE and 1 mM GTP for 30 min at 37°C. All samples were then incubated with 1 μg HCM and 100 μM SAM for 30 min at 37°C. C: capped transcripts recovered with anti-2,2,7-trimethylguanosine beads. U: uncapped transcripts found in supernatant. Lengths of limiting-CTP transcripts are indicated; brackets indicate sizes of capped and uncapped transcripts. 9% Urea-PAGE.
**Discussion**

Based on my occupancy data for H3K4me1, H3K4me3, and transcriptional machinery, I propose a model where H3K4me1 is deposited near sites of Pol II transcription and is “upgraded” to H3K4me3 as a result of regular productive elongation. H3K4 methylation is carried out by Set1A and Set1B in humans (Kim et al., 2009; Zhu et al., 2005) and their activities are closely linked with the PAF1 complex (PAF1C), which is widely accepted to facilitate productive elongation (Crisucci and Arndt, 2011). Two conflicting views were recently published on the role of PAF1 in the transition into productive elongation genome-wide (Chen et al., 2015; Yu et al., 2015). In one, PAF1 was described to maintain Pol II pausing (Chen et al., 2015); the other stated PAF1 facilitated the release of paused Pol II into productive elongation (Yu et al., 2015). Despite these differing conclusions, both studies contained ChIP-Seq of PAF1 complex members and, in agreement with a previous CTR9 dataset in mouse embryonic stem cells (Rahl et al., 2010), found PAF1C throughout the coding regions and past the 3′ transcription end sites of active genes. Notably, PAF1C subunits were described or shown to overlap with the TSS of the most Pol II-occupied genes (Chen et al., 2015; Yu et al., 2015). While this would explain the H3K4me1/3 occupancies I describe around promoters, the PAF1C was not observed over enhancer regions in their datasets. This could be due to the high backgrounds present in their datasets or the lack of enhancer activity in the cell types used. This question could be answered directly by knocking down PAF1 and looking for changes in H3K4me1 and H3K4me3 ChIP-Seq occupancy. If PAF1 and Set1 do not influence enhancer regions, H3K4me1 could be deposited by MLL proteins, which have a minimal role near promoters but could act near enhancers (Wu et al., 2008).

In agreement with previous H2A.Z studies, my ChIP-Seq data showed that H2A.Z is generally in promoter and enhancer regions co-occupied by transcriptional machinery (Chen et al., 2013; Yukawa et al., 2014). H2A.Z was also frequently present in promoter-proximal nucleosomes, and this enrichment has been previously shown to be inversely correlated with Pol II stalling and MNase accessibility (Weber et al., 2014). However, H2A.Z appeared to weakly anticorrelate with productive elongation and was absent downstream of the most transcribed genes. Unlike promoters, where H2A.Z
occupancy was limited to promoter-proximal nucleosomes, H2A.Z substitution was broadly detected across enhancer regions, the majority of which were depleted of H3K4me3 and other indicators of productive elongation. My data does not support a model where H2A.Z substitution is meant to facilitate productive elongation through chromatin. Because H2A.Z and H3.3 are frequently found together in nucleosomes at promoters and enhancers (Goldberg et al., 2010; Jin et al., 2009; Yukawa et al., 2014) and H3.3 offsets the stabilization effects of H2A.Z on chromatin (Chen et al., 2013; Jin and Felsenfeld, 2007), it is possible that H2A.Z/H3.3 double-substitution “loosens” chromatin around promoters and enhancers and in this model, these nucleosomes would act as easily removable placeholders for future transcription by Pol II. H2A.Z/H3.3 might reserve genomic regions for future Pol II occupancy and, unlike more stable histone barriers, could be more readily removed by chaperones (Gursoy-Yuzugullu et al., 2015; Jeronimo et al., 2015) once productive elongation resumes. Additional research is required to determine if H2A.Z directly influences Pol II elongation and, if so, if it facilitates or hinders transcription.

In addition to identifying the inducible *HEXIM1* promoter (Liu et al., 2014), my RNA-Seq data suggests that *HEXIM1* expression could be regulated in part by transcriptional interference. Because transcriptional interference would prevent the establishment of promoter-proximal paused polymerases at the sense *HEXIM1* promoter, the presence of Pol II, DSIF, and NELF and the downstream accumulation of H3K4me3 indicate that convergent transcription from *HEXIM2* is either infrequent, or stops before the *HEXIM1* promoter. Noncoding antisense transcription has been previously reported to result in nuclear RNA interference (RNAi) over a number of genes (Faghihi and Wahlestedt, 2009; Haussecker and Proudfoot, 2005; Janowski et al., 2005; Morris et al., 2004). Dicer, while found primarily in the cytoplasm, can localize to the nucleus (Ando et al., 2011; Doyle et al., 2013) and interact directly with elongation complexes in a dsRNA-dependent manner (White et al., 2014). Importantly, Dicer suppresses the accumulation of dsRNA products, which prevents activation of the interferon pathway and apoptosis (White et al., 2014). In addition to the potential destruction of sense *HEXIM1* transcripts, RNAi machinery can also modulate DNA methylation and histone modification to potentially silence this region (Holoch and
Moazed, 2015; Kalantari et al., 2016). Potential strategies for establishing the role of RNAi at *HEXIM1* include use of shRNAs against the antisense *HEXIM2* lncRNA transcript or, alternatively, deletion of the antisense *HEXIM2* promoter by CRISPR. Study of the *HEXIM2* promoter before and after *HEXIM1* induction could identify additional regulatory mechanisms.

My analyses and other results (Guo et al., 2014a) support a model for Myc function where the transcription machinery, rather than DNA sequence elements, plays the major role in recruiting the Myc-Max heterodimer to genomic sites. Such protein-protein interactions could occur through the unstructured N-terminal transcription activation domain of Myc (Nair and Burley, 2003). Myc-Max was shown to readily bind non-specific DNA sequences—the affinity of Myc-Max to specific DNA is only about 100-fold stronger than non-specific DNA by EMSA (Guo et al., 2014a) and 56-fold stronger by protein binding microarray—and this interaction with DNA likely stabilizes any greater Myc-Max complex formed with Pol II. Mediator probably participates in such a complex and interacts with both Pol II (Allen and Taatjes, 2015; Plaschka et al., 2015) and Myc (Liu et al., 2008). This model could apply to other transcription factors. For example, significant overlap can be seen between the occupancies of Myc and transcription factors AP-1 and AP-2 (Consortium, 2012). Even STAT5A, which specifically occupied a specific DNA sequence element over the HIV promoter after HODHBt treatment, appeared to colocalize with Pol II genome-wide (data not shown), though that ChIP was performed with excessive crosslinking and a small number of cells. Overall, the genomic landscape may be too complex for sequence specificity to solely govern the occupancies and activities of transcription factors.

Although my ChIP-Seq data demonstrated that STAT5A is specifically recruited to the HIV promoter in naturally-infected primary T cells, the contribution of promoter-proximal pausing to HIV latency in these cells remains unclear. Promoter-proximal pausing was visible in a transformed cell model of HIV infection (Jadlowsky et al., 2014) but because ChIP-Seq is the sum of a snapshot in millions of cells, the uniform nature of HIV integration and latency in these cells resulted in an artificially clear picture. Due to the amount of cell death after HIV infection, I was only able to use 1.2-6 million cells per dataset; increasing the number of cells per ChIP could reduce non-
specific background signal but this may not be feasible. Transcriptional interference (Han et al., 2008; Lenasi et al., 2008) could also result in my observed distribution of Pol II. ChIP-Seq against NELF before and after flavopiridol treatment, which inhibits P-TEFb (Chao et al., 2000; Chao and Price, 2001) and prevents transcription (Cheng et al., 2012), could establish the degree to which promoter-proximal pausing occurs on the HIV promoter in the presence or absence of transcriptional interference from nearby genes. This experiment would potentially validate a recent finding where NELF-E bound specifically to the nascent transcript of HIV and influenced pausing over the LTR (Pagano et al., 2014). PRO-Seq could also be used to directly measure the ratio between antisense transcripts from interfering elongation and sense transcripts from pausing. If performed in collaboration with Alberto Bosque and Vicente Planelles, these experiments would answer lingering questions about how latency is maintained in naturally-infected T cells.

Because HODHBt prevents SUMOylation of the active, phosphorylated form of STAT5, it can only work in cells with functional STAT5. Of note, the majority of activated STAT5 present in CD4-positive T cells isolated from HIV-infected patients is truncated (Bovolenta et al., 1999) and this form acts as a dominant-negative inhibitor of HIV transcription (Crotti et al., 2007). In the cultured central-memory-like T cells used in my studies, only full length STAT5 was detected. However, reactivation of latent HIV was still achieved in latently-infected T cells isolated from HIV patients and these cells should have truncated STAT5 (data not collected). Additional experiments could be performed to determine if the degree of HIV induction after HODHBt treatment and the amount of truncated STAT5 are correlated. Circulating memory T cells are not the only reservoir of latent HIV; persistent viruses have been detected in both lymphoid and myeloid lineages (Chun et al., 2015). Because STAT5 regulates proliferation and differentiation in myeloid cells (Ilaria et al., 1999), stabilization of phosphorylated STAT5 with HODHBt could have unwanted side-effects in patients. The ability of HODHBt to reactivate HIV in these cells remains undetermined. Ultimately, multiple drugs may be required to adequately target and eliminate HIV reservoirs in patients. Elimination of these reservoirs could enable infected patients to cease antiretroviral therapy without fear of future onset of AIDS, effectively preventing the disease.
CHAPTER 3: CDK7 LINKS CAPPING AND ELONGATION CONTROL

This chapter focuses on the effects of THZ1, a covalent inhibitor of cyclin-dependent kinase 7 (Cdk7), on co-transcriptional capping, promoter proximal pausing, and productive elongation. Portions of this chapter were previously published (Nilson et al., 2015) and are reused with permission.

Introduction

Cdk7, Cyclin H, and Mat1 comprise a kinase module in the Pol II initiation factor TFIIH (Egly and Coin, 2011) that is responsible for phosphorylation of the CTD of the large subunit of human Pol II during initiation. The CTD contains 52 repeats of a heptapeptide with the consensus sequence $Y₁S₂P₃T₄S₅P₆S₇$ and Cdk7 phosphorylates Ser5 and Ser7 residues, the first of several CTD phosphorylation and dephosphorylation events during the transcription cycle (Heidemann et al., 2013). CTD phosphorylation is thought to be important in the recruitment of RNA processing factors and histone modification enzymes (Heidemann et al., 2013). In mammals, Cdk7 is also the major cyclin-dependent kinase activating kinase and is involved in regulating the cell cycle (Fisher, 2012).

A prominent feature of transcription in metazoans is promoter-proximal pausing of Pol II (Guo and Price, 2013). Pausing is induced by the DRB sensitivity inducing factor, DSIF, in conjunction with the negative elongation factor NELF (Yamaguchi et al., 2013) and these factors accumulate with Pol II downstream of the transcription start site of the majority of mammalian genes (Liu et al., 2014; Rahl et al., 2010). Gdown1 has also been implicated in stable pausing of Pol II near promoters (Cheng et al., 2012). To transition into productive elongation, the Cdk9 kinase subunit of P-TEFb must phosphorylate both DSIF, which remains associated during elongation, and NELF, which is released (Guo and Price, 2013). Cdk9 (Marshall et al., 1996) and Cdk12 (Bartkowiak et al., 2010; Bowman and Kelly, 2014) have been implicated in phosphorylation of CTD Ser2 during the transition into productive elongation and at the 3’ ends of genes.

Capping occurs co-transcriptionally and is the first of many processing steps required to generate a functional mRNA (Bentley, 2014). Initially, an RNA triphosphatase removes the terminal phosphate and then RNA guanylyltransferase
mediates the 5′–5′ addition of GMP. Cap methylation occurs in a separate step and the final m⁷G cap influences downstream events. Guanylylation of nascent transcripts can occur as soon as the RNA emerges from Pol II (Coppola et al., 1983; Moteki and Price, 2002; Rasmussen and Lis, 1993) and human capping enzyme (HCE) is 4 to 5 orders of magnitude more efficient when the RNA substrate is associated with an elongation complex (Moteki and Price, 2002). This is aided in part by capping enzyme association with the phosphorylated CTD (Chiu et al., 2002; Cho et al., 1997; McCracken et al., 1997a; Yue et al., 1997), although the site of this interaction is not conserved between yeast and mammalian capping enzymes (Ghosh et al., 2011). In fission yeast, an otherwise lethal Ser5Ala CTD mutation was rescued by fusing mouse capping enzyme to the CTD (Schwer and Shuman, 2011). In mammals, CTD phosphorylation has a 4-fold effect on capping of soluble RNA (Ho and Shuman, 1999) or nascent Pol II transcripts (Mandal et al., 2004; Moteki and Price, 2002). Capping enzymes have also been shown to interact with the Spt5 subunit of DSIF in yeast (Doamekpor et al., 2014; Doamekpor et al., 2015; Pei and Shuman, 2002; Schneider et al., 2010) and human DSIF causes a 2 to 5-fold stimulation of co-transcriptional capping (Mandal et al., 2004; Wen and Shatkin, 1999). Additionally, wild type or capping-defective HCE relieved the negative influence of NELF during transcription in vitro suggesting a competition of HCE and pausing factors for Pol II (Mandal et al., 2004). This network of interactions among the capping enzyme, the Pol II CTD, and DSIF could play a role in controlling capping and elongation in mammals.

Inhibition of Cdk7 has broad effects on transcription but the mechanisms involved remain poorly understood. Early in vitro studies gave conflicting results concerning the role of Cdk7 in transcription (Akoulitchev et al., 1995; Tirode et al., 1999). Inhibition of the Cdk7 homolog Kin28 in budding yeast reduced mRNA capping in two studies. In one, global transcription was unaffected (Kanin et al., 2007) and in the other, mRNA stability was reduced (Hong et al., 2009). Other studies showed increased Pol II promoter occupancy genome-wide (Bataille et al., 2012; Wong et al., 2014). Investigations in higher eukaryotes showed decreases in promoter-proximal pausing (Chipumuro et al., 2014; Glover-Cutter et al., 2009; Kelso et al., 2014; Kwiatkowski et al., 2014; Larochelle et al., 2012; Schwartz et al., 2003). This conflict is likely due to
differences in elongation control machinery. Two recent reports described reductions in DSIF and increases in TFIIE near promoters in human cells as a result of Cdk7 inhibition (Kelso et al., 2014; Larochelle et al., 2012), but while Larochelle et al. described a concurrent loss of NELF ChIP signal, Kelso et al. found that NELF association with transcription complexes in vitro was unaffected. Curiously, Kelso et al. also showed that inhibition of Cdk7 reduced in vitro transcription but only affected 2% of expressed genes by microarray.

A newly developed covalent inhibitor of Cdk7, THZ1, negatively impacts gene expression and proliferation in cancer cell lines and animal models (Chipumuro et al., 2014; Christensen et al., 2014; Kwiatkowski et al., 2014; Wang et al., 2015). To elucidate the mechanism of THZ1 action, I used an in vitro system with HeLa nuclear extract to examine how the compound affects transcription and associated events. By studying the effects of THZ1 on Pol II initiation, pausing, and productive elongation, I discovered that THZ1 causes a cascade of defects that ultimately impair capping and elongation factor recruitment. Importantly, these results can explain the effects of THZ1 seen in cells. This research will aid evaluation of THZ1 as a potential cancer therapy and serve as a template for future studies of other compounds.

**Materials and Methods**

*In vitro transcription*

The use of soluble and paramagnetic bead-immobilized CMV templates and HeLa nuclear extract (HNE) to study transcription in vitro was described previously (Adamson et al., 2003; Cheng and Price, 2007, 2009). All steps were performed at room temperature (RT). Template DNA from CMV -800 bp to +508 bp was incubated for 30 min with 1 μl/rxn HNE in the presence of 60 mM KCl, 5 mM MgCl$_2$, 20 mM HEPES pH 7.6, 1 mM DTT, and 0.5 U/μl SUPERase-In. 1 mM THZ1 (Kwiatkowski et al., 2014) and flavopiridol (Chao et al., 2000) stocks in DMSO were diluted to 10 μM in H$_2$O immediately before use. In Figure 33, 1 mM DTT was replaced with 3 mM H$_2$O$_2$ as indicated (2 mM final during the chase). Initiation was accomplished with a 30 or 45 s pulse containing 60 mM KCl, 5 mM MgCl$_2$, 20 mM HEPES pH 7.6, 0.21 μM α-$^32$P-CTP, and 500 μM ATP/UTP/GTP (limiting CTP) or 500 μM ATP/GTP 0.1 μM UTP (limiting UTP/CTP). Elongation complexes were either stopped by addition of EDTA to 20 mM or
chased for 3 min with 500 μM CTP. Before the chase, immobilized complexes were isolated as indicated with high salt wash (1.6 M KCl, 20 mM HEPES pH 7.6, 1 mM DTT, and 0.02% Tween20) to remove associated factors, or low salt wash (60 mM KCl, 20 mM HEPES pH 7.6, 1 mM DTT, 0.2 mg/ml BSA, and 0.02% Tween20) to allow retention of associated factors. Labeled transcripts were extracted with phenol, precipitated with 95% ethanol and 500 mM NH₄C₂H₃O₂, separated on denaturing RNA gels (6 M urea, 1X TBE, and 6 or 9% 37.5:1 acrylamide:bis-acrylamide), scanned using a Fujifilm Typhoon FLA-7000 phosphorimager, and analyzed using Fujifilm MultiGauge v3 software.

**Kinase assay**

Kinase assay conditions were adapted from (Marshall et al., 1996). 10 mM THZ1 and flavopiridol were serial diluted in 1% DMSO and preincubated for 10 min in 15 μl reactions containing 1 pmole recombinant DSIF (Renner et al., 2001), 0.04 pmoles recombinant P-TEFb (Cheng and Price, 2007), 25 mM HEPES pH 7.2, 50 mM KCl, 6.67 mM MgCl₂, 1.33 mM DTT, 0.067 mg/ml BSA, and 0.02% Triton X-100. After 10 min, reactions were completed to 20 μl and incubated for 10 min with the following buffer conditions: 25 mM HEPES pH 7.2, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, 0.05 mg/ml BSA, 0.02% Triton X-100, 30 μM cold ATP, and 2.5 μCi γ-³²P-ATP (PerkinElmer BLU002Z500UC). Reactions were stopped by addition of 5 μl 5X protein loading buffer (20% Ficoll, 10% SDS, 50 mM EDTA, 50 mM DTT, and 50 mM Tris pH 7.6), denatured at 95° for 5 min, separated by 9% SDS-PAGE electrophoresis (1X Tris-glycine with 0.1% SDS, 9% 37.5:1 acrylamide:bis-acrylamide resolving gel, and 4% 19:1 acrylamide:bis-acrylamide stacking gel), and analyzed by silver staining and phosphorimaging.

**Factor add-backs**

Recombinant human capping enzyme (HCE) was purified as described previously (Moteki and Price, 2002). The following steps were performed at RT. Serial dilutions of HCE in low salt wash were used for Figures 29-32. High or low salt isolated elongation complexes were incubated for 1 min (Figure 29) or 3 min (Figure 30-31) with equal volumes of reaction buffers containing 60 mM KCl, 5 mM MgCl₂, 20 mM HEPES pH 7.6, 1 mM DTT, 0.2 mg/ml BSA, 500 μM GTP, and indicated amounts of HCE. Capping
was stopped with a solution of 1% sarkosyl, 100 mM NaCl, 50 mM Tris pH 7.6, and 20 mM EDTA immediately before phenol extraction.

Recombinant DSIF (Renner et al., 2001), NELF (Renner et al., 2001), and Gdown1 (Cheng et al., 2012) were purified as described previously. The following steps were performed at RT and all factors were diluted 10-fold in low salt wash immediately before use. High or low salt isolated complexes were incubated for 5 min prior to chase with equal volumes of reaction buffers containing 60 mM KCl, 20 mM HEPES pH 7.6, 1 mM DTT, 0.2 mg/ml BSA, 0.5 U/μl SUPERase-In, and 1) no additional factors, 2) 0.3 pmoles DSIF and 0.6 pmoles NELF, 3) 0.6 pmoles NELF, or 4) 1 pmoles Gdown1. All add-back reactions were stopped with a solution of 1% sarkosyl, 100 mM NaCl, 50 mM Tris pH 7.6, and 20 mM EDTA immediately before phenol extraction.

**Transcript cap status determination**

Cap status determination using recombinant human cap methyltransferase (HCM) was described previously (Moteki and Price, 2002). For Figures 29, 33, and 37, anti-2,2,7-trimethylguanosine agarose beads (Calbiochem NA02A) which recognize m⁷G-capped RNAs were blocked for 30 min at RT in blocking buffer (25 mM Tris pH 7.6, 5 mM Mg(C₂H₃O₂)₂, 25 mM KC₂H₃O₂, 25 mM NaCl, 1 mM DTT, 0.02% Tween20, 0.05 mg/ml Torula yeast RNA (Sigma R6625), 0.2 mg/ml BSA, and 0.02 U/μl SUPERase-In), washed three times with washing buffer (25 mM Tris pH 7.6, 5 mM Mg(C₂H₃O₂)₂, 75 mM KC₂H₃O₂, 75 mM NaCl, 1 mM DTT, 0.1% Tween20, and 0.02 U/μl SUPERase-In), rinsed three times with binding buffer (blocking buffer without BSA), and resuspended in binding buffer as a 20% slurry. Additionally, anti-2,2,7-trimethylguanosine ascites fluid (Calbiochem CS214155) was diluted 1:10 in 35 mM KCl HGKEDP (25 mM HEPES pH 7.6, 15% glycerol, 35 mM KCl, 0.1 mM EDTA, 1 mM DTT, and 0.1% isopropanol-saturated PMSF) containing 0.5% Triton X-100 and fractionated with an 80 ml gradient (50 mM to 1 M KCl HGKEDP) over a Mono Q HR 10/10 column. Pure IgG eluted around 100 mM KCl and antibody-containing fractions were pooled. For Figure 39, Mono Q-purified anti-2,2,7-trimethylguanosine antibody was bound to Protein G Sepharose 4B Fast Flow beads (equilibrated in blocking buffer lacking Torula yeast RNA) for 1 h at 4° with rotation. Antibody-beads were washed once with RIPA buffer (20 mM Tris pH 7.6, 150 mM KC₂H₃O₂, 1% Triton X-100, 0.2% sodium deoxycholate, 0.1% SDS, and 1 mM EDTA),
once with LiCl buffer (20 mM Tris pH 7.6, 400 mM LiCl, 0.02% Tween20, and 1 mM EDTA), and three times with binding buffer lacking Torula yeast RNA.

Radiolabeled transcripts were phenol extracted, ethanol precipitated, and resuspended in 10 μl/rxn 25 mM Tris pH 7.6 with 0.5 U/μl SUPERase-In. Half of each sample was set aside for analysis of total transcripts. The other half was incubated for 15 min at 37° with an equal volume of 25 mM Tris pH 7.6, 10 mM Mg(C₂H₃O₂)₂, 200 μM SAM, 1 mM DTT, 0.2 mg/ml BSA, and 30 ng/μl HCM to methylate existing RNA caps. After methylation, 50 μl/rxn 20% antibody bead slurry was added to each sample and incubated with rotation for 2-4 h at RT. Samples were then washed three times with 70 μl washing buffer; unbound supernatants were saved and pooled. Total, bound, and unbound fractions were brought to equal volumes in washing buffer, spiked with glycogen, and phenol extracted, precipitated, and analyzed as described above.

Nuclear run-on

Nuclei isolation conditions were adapted from (Chao and Price, 2001). HeLa S3 cells were grown at 37° and 5% CO₂ to 80% confluency in T75 flasks in 30 ml DMEM (Gibco 11965-092) supplemented with 10% FBS. One hour before harvesting, 10 ml of media was removed, spiked with 1:1,000 final volume of DMSO containing 1 mM of flavopiridol or THZ1 as indicated (1 μM final), and re-added to cells. After compound incubation, adherent cells were aspirated, rinsed with 20 ml ice-cold PBS, incubated for 5 min on ice with 20 ml swelling buffer (10 mM Tris pH 7.6, 2 mM Mg(C₂H₃O₂)₂, and 3 mM CaCl₂), scraped, and pelleted at 500 × g for 5 min. Cell pellets were resuspended in 15 ml ice-cold lysis buffer (10 mM Tris pH 7.6, 320 mM sucrose, 0.5% Triton X-100, 2 mM Mg(C₂H₃O₂)₂, 3 mM CaCl₂, 1 mM DTT, 0.004 U/μl SUPERase-In, and 0.1% isopropanol-saturated PMSF) and Dounce homogenized. Cell lysis was monitored by phase-contrast microscopy. Once lysed, cells were pelleted at 1,200 × g for 5 min and thoroughly resuspended in a mixture of 1 ml lysis buffer and 2 ml sucrose cushion (10 mM Tris pH 7.6, 1.9 M sucrose, 5 mM Mg(C₂H₃O₂)₂, 1 mM DTT, 0.004 U/μl SUPERase-In, and 0.1% isopropanol-saturated PMSF). This homogenate was carefully layered over a 2 ml sucrose cushion and spun at 30,000 × g for 45 min. Nuclei were resuspended in 1 ml storage buffer (10 mM Tris pH 7.6, 25% glycerol, 5 mM Mg(C₂H₃O₂)₂, and 5 mM DTT),
pelleted at 1,200 × g for 5 min, and resuspended with storage buffer to about 5 × 10^7 nuclei/ml before storage in aliquots at -80°.

Nuclear run-on nucleotide conditions were adapted from published studies (Core et al., 2012; Core et al., 2008). 5 × 10^5 DMSO-, flavopiridol-, or THZ1-treated HeLa S3 nuclei were first diluted to 25 μl in storage buffer, and then to 50 μl with 10 mM Tris pH 7.6, 5 mM Mg(C_2H_3O_2)_2, 5 mM DTT, and 0.25 U/μl SUPERase-In. Each sample was equally divided and incubated for 4 min at 30° with an equal volume of reaction buffer containing 10 mM Tris pH 7.6, 1% sarkosyl, 300 mM KC_2H_3O_2, 5 mM Mg(C_2H_3O_2)_2, 5 mM DTT, 0.5 U/μl SUPERase-In, 500 μM ATP/UTP/GTP, 2 μM cold CTP, 0.33 μM α-³²P-CTP, and 4 μg/ml α-amanitin as indicated. Reactions were stopped by addition of EDTA to 20 mM and labeled transcripts were extracted with Trizol LS, precipitated with 95% ethanol and 500 mM NH_4C_2H_3O_2, and assayed for capping as described above.

Nuclear walk-on

Nuclei isolation was performed as described above with the following changes. HeLa cells were grown to 5 × 10^6 cells/ml in spinner flasks at 37° and 5% CO_2 in SMEM supplemented with 10% FBS and 2 mM L-Glutamine. After compound incubation, suspension cells were pelleted at 500 × g for 10 min and media was decanted. Lysis, sucrose cushion sedimentation, and storage conditions were unchanged.

2.5 × 10^6 nuclei were diluted to 50 μl in storage buffer lacking glycerol. Two 20 μl aliquots were incubated for 3 min at 30° with an equal volume of reaction buffer containing 10 mM Tris pH 7.6, 1% sarkosyl, 300 mM KC_2H_3O_2, 5 mM Mg(C_2H_3O_2)_2, 5 mM DTT, 0.5 U/μl SUPERase-In, 0.33 μM α-³²P-CTP, and 4 μg/ml α-amanitin as indicated. Reactions were stopped, precipitated, and assayed for capping as described above.

Results

Effects of THZ1 on transcription in vitro

To gain mechanistic insight into the effects of THZ1, I used an in vitro system that reconstitutes initiation, pausing, productive elongation, and coupled RNA processing (Adamson et al., 2003). In this system, HeLa nuclear extract and a CMV-promoter-driven template are preincubated for 30 minutes to allow formation of preinitiation complexes (PICs). Initiation is then accomplished in a 30 second pulse
with limiting α-³²P-CTP. When added in increasing amounts to the preincubation step, THZ1 changed the pattern but not the amount of transcripts generated during the pulse and these changes plateaued at 1 μM (Nilson et al., 2015). While THZ1 covalently inhibits Cdk7, it can also non-covalently inhibit Cdk12 and Cdk13 at much higher concentrations (Kwiatkowski et al., 2014). Additionally, a non-covalent THZ1 analog had no effect on Cdk7 in the concentration range used in this study (Kwiatkowski et al., 2014). As covalent interactions take longer to form than non-covalent interactions, preincubation should be required if the effects of THZ1 are due to Cdk7 inhibition.

To determine the timing of THZ1 action, I preincubated PICs with 1 μM THZ1 for increasing amounts of time (Figure 27). In the absence of THZ1, a typical distribution of transcripts was generated by a 30 second limiting-CTP pulse (elongation complexes should stop after generating transcripts 16, 21, 25, 31, 49, 54, and 56 nt in length). Transfer RNAs (tRNA) will also incorporate CTP due to CCA addition in the presence of extract and should be ignored. Doublet RNA bands were observed at all sizes except 16 nt; I showed in Figure 26 that these doublets are from capping. Because transcripts shorter than 20 nt do not extend beyond the body of the polymerase, they cannot be capped co-transcriptionally (Coppola et al., 1983). When THZ1 was added during the pulse, no changes in the pattern were observed. One minute of THZ1 preincubation, however, resulted in a slight increase of the more rapidly migrating (uncapped) U25 and A31 transcripts. When THZ1 acted for 3 minutes before the pulse, capped transcripts decreased while uncapped transcripts increased. This change in migration plateaued after 10 minutes of THZ1 treatment; all transcripts including G21 were mostly uncapped. Because THZ1 required at least 3 minutes of preincubation to influence capping, these effects are likely related to covalent inhibition of Cdk7.

1 μM THZ1 also inhibited productive elongation (Nilson et al., 2015). To determine if THZ1 inhibits the Cdk9 kinase subunit of P-TEFb, a factor required for the transition into productive elongation, I performed kinase assays with P-TEFb, DSIF as substrate, 30 μM cold ATP with 2.5 μCi γ-³²P-ATP, and increasing concentrations of THZ1 or flavopiridol, a potent P-TEFb inhibitor (Figure 28). As expected, low concentrations of flavopiridol reduced phosphorylation of Spt5, a subunit of DSIF. P-TEFb was 79% inhibited at 0.01 μM and completely inhibited at 0.1 μM. THZ1 had no
HeLa nuclear extract and template DNA were preincubated for 30 min and pulsed for 30 s with limiting $\alpha$-$^{32}$P-CTP. THZ1 was introduced as indicated to give a final pulse concentration of 1 μM. In the presence of nuclear extract, tRNAs incorporate $\alpha$-$^{32}$P-CTP independently of transcription. Lengths of limiting-CTP transcripts are indicated and their stop sites are underlined in the sequence below. Brackets indicate sizes of capped and uncapped transcripts. 9% Urea-PAGE.

effect at these concentrations and only had a partial effect at 1 μM. Because Cdk9 lacks a cysteine near its ATP binding cleft for THZ1 to link with covalently (Kwiatkowski et al., 2014), any interaction with P-TEFb is likely to be non-covalent in nature. As in vitro transcription reactions contain significantly more ATP (500 μM) than this kinase assay (30 μM), the effects of THZ1 on P-TEFb would be further reduced by 17-fold. This result suggests that the effects of 1 μM THZ1 are unlikely to be from P-TEFb inhibition.

THZ1 inhibits early, efficient capping

Next, I sought to clarify the effects of THZ1 on capping. As diagrammed in Figure 29, complexes were pulsed for 45 seconds with limiting CTP to generate a diverse pattern of transcripts whose cap status was assessed. Capped transcripts have
Figure 28. Effects of THZ1 and flavopiridol on P-TEFb kinase activity

1 pmole DSIF and 0.04 pmole P-TEFb were incubated for 10 min with THZ1 or flavopiridol and 10 min with 30 μM cold ATP with 2.5 μCi γ-³²P-ATP. Final concentrations of THZ1 and flavopiridol are indicated. Label incorporation into the Spt5 subunit of DSIF was quantified and plotted in the lower panel. Note that in vitro transcription reactions instead use 500 μM ATP and non-covalent THZ1 activity will be reduced 17-fold. 9% SDS-PAGE followed by phosphorimaging (top) and silver staining (bottom).

one less negative charge and one additional nucleoside, thus lowering their mobility in denaturing gels. To confirm their cap status, transcripts were isolated, treated with the human cap methyltransferase and SAM to mature existing caps, and then captured with anti-2,2,7-trimethylguanosine beads. G16 was uncapped in both reactions because its 5′ end is still protected within Pol II. In the control reaction, transcripts ending in G21 were partially capped and all larger transcripts were almost entirely capped as indicated by their mobilities and their retention by the anti-cap beads. As seen before in Figure 27, 1 μM THZ1 treatment abolished capping of G21 transcripts and reduced the capping of longer transcripts.

To determine if the THZ1-dependent loss of capping of G21 and longer transcripts was due to altered CTD phosphorylation or associated factors, I high salt
Figure 29. THZ1-treated elongation complexes respond differently to HCE
Cap status determination after 30 min preincubation ± THZ1, 45 s limiting CTP pulse, high salt wash, and incubation with or without 1 pmole HCE for 1 min. Final concentration during the pulse was 1 μM for THZ1. Below: transcript sequence with limiting-CTP stops underlined. C: capped transcripts recovered with anti-2,2,7-trimethylguanosine beads. U: uncapped transcripts found in supernatant. 9% Urea-PAGE. washed pulsed complexes and incubated with recombinant HCE and GTP for 1 minute. G16 remained uncapped but G21 was now completely capped for both control and THZ1 complexes (Figure 29). This suggests that the inefficient capping of G21 observed during pulse reactions is modulated by a factor that was removed by high salt wash. Longer nascent transcripts (U25, A31, G49-56) were already capped in the control elongation complexes but these transcripts, which were poorly capped during the THZ1 pulse, were still only partially capped by recombinant HCE and the longer transcripts were capped more efficiently than U25. These data demonstrate that 21 nucleotides is intrinsically the preferred length for efficient capping, but capping at this position may normally be down-modulated by a factor associated with the elongation complex. It seems likely this factor could be involved in the THZ1-dependent increase in inhibition of capping of G21.
Figure 30. Activity of HCE added to high salt washed elongation complexes ± THZ1

High salt washed elongation complexes generated as in Figure 29 were incubated for 3 min with indicated amounts of HCE.

To further examine recruitment of the capping enzyme, I titrated HCE over a range which spanned the concentration of high salt washed Pol II complexes. About 30 femtomoles of elongation complexes were generated in the presence or absence of THZ1, high salt washed, and incubated for 3 minutes with 3 to 300 femtomoles of recombinant HCE. Substoichiometric addition of capping enzyme to control complexes was able to complete G21 capping (Figure 30). Approximately 3-fold more HCE was needed to completely cap G21 transcripts associated with THZ1 complexes, which lack CTD phosphorylation (Nilson et al., 2015). This is quantitatively similar to previous observations where CTD phosphorylation had only a 4-fold effect on coupled capping activity (Moteki and Price, 2002). As seen in Figure 29 for THZ1-treated complexes, HCE efficiency was transcript length-dependent. These results support the idea that capping is most efficient at the G21 position in the absence of additional factors.

To assess the influence of factors associated with the elongation complex on the activity of capping enzyme, I next added increasing amounts of HCE to low salt washed complexes. About one third of the G21 transcripts in control complexes were resistant to levels of HCE 1,000 times greater than that needed to completely cap high salt washed complexes (Figure 31). These results suggest that a factor that inhibits capping of G21 was retained by a fraction of the complexes after low salt wash. THZ1 complexes
Figure 31. Activity of HCE added to low salt washed elongation complexes ± THZ1

Complexes generated as in Figure 29 except with a low salt wash step were incubated for 3 min with indicated amounts of HCE.

behaved in a similar manner with about a third of the complexes being resistant to the highest level of HCE tested. Because THZ1 did not increase the association of this capping inhibitor, additional factors or processes must be involved in the THZ1-mediated inhibition of capping.

To examine the efficiency of capping on complexes that were not washed, I titrated capping enzyme into the preinitiation complex assembly reactions. Capping of G21 during the pulse reaction for control and THZ1-treated complexes was resistant to all but the highest concentrations of HCE (Figure 32). Based on the amount of capping enzyme needed, the efficiency of capping of G21 during the pulse was again about 3 orders of magnitude less efficient than observed with high salt washed complexes. THZ1-treated complexes were more responsive to capping at U25 and A31 than at position G21, suggesting the inhibitor functions mainly at G21. It is interesting that a fraction of the G21 transcripts were capped with the very low levels of HCE in extract, but the rest were resistant to orders of magnitude increases in HCE. This suggests that loss of the capping inhibitor at G21 may be part of an ordered exchange of factors that normally occurs during progression to the promoter proximal paused state and this transition is blocked by THZ1.
Figure 32. Activity of HCE added to preinitiation complexes ± THZ1
Complexes generated as in Figure 29 except indicated amounts of HCE were added 20 min before initiation of transcription by pulse.

Taken together, these results indicate that in the absence of additional factors, nascent transcripts are capped with high efficiency just as they emerge from the RNA exit channel at G21. However, capping at G21 is several orders of magnitude less efficient in the presence of extract which strongly suggests there is a factor that inhibits capping at this site. If capping at this “sweet spot” is missed or blocked, such as in the presence of THZ1, HCE must wait for transcripts to become longer before capping can occur and this capping is less efficient. The mechanism of THZ1 inhibition of capping at G21 likely involves an enhancement of the function of the capping inhibitor but how that is achieved is not yet clear. Loss of CTD phosphorylation may be involved in this mechanism but its small direct effect on capping cannot explain the large effect seen in the presence extract without evoking an additional unknown factor.

THZ1 effects on pausing are due to defective recruitment of DSIF and NELF, not impaired capping

Although THZ1 was shown to inhibit CTD phosphorylation, capping, pausing, and productive elongation (Nilson et al., 2015), it was ambiguous if these effects were functionally coupled or if they arose from parallel pathways that could have common intermediates. To determine if the loss of capping was causative or merely correlative with effects on pausing, I performed pulse/chase experiments in the presence of
flavopiridol to block productive elongation and examined the patterns of total, capped and uncapped transcripts (Figure 33A). To aid interpretation of these results, I generated profiles of counts incorporated from about 90-550 nt (Figure 33B-D). In the absence of P-TEFb activity, control complexes exhibited poor elongation as evidenced by the accumulation of transcripts less than 300 nt in length, and the relative absence of transcripts longer than 300 nt (Figure 33B). Not only were control complexes paused, essentially all of their transcripts were capped (Figure 33C-D). Complexes treated with THZ1 generated longer transcripts overall (Figure 33B) but interestingly, half of the nascent transcripts were capped and appeared to be associated with paused Pol II (Figure 33C); only uncapped THZ1 transcripts were significantly longer (Figure 33D). This result is consistent with capping being required for pausing. Interestingly, 2 mM H$_2$O$_2$ inhibited capping without affecting pausing. This effect was best visualized after normalization to account for a slight inhibition of initiation by H$_2$O$_2$ (Figure 33E).
Figure 34. Effects of H$_2$O$_2$ on HCE added to high salt washed elongation complexes
HCE was preincubated for at least 5 min with increasing amounts of H$_2$O$_2$ before addback to high salt washed elongation complexes as in Figure 30. Figures 30 and 34 were run on the same gel and the first lane is shared. 9% Urea-PAGE.

Although I do not know how H$_2$O$_2$ blocks capping, the same concentration of H$_2$O$_2$ did not affect Pol II CTD phosphorylation (Nilson et al., 2015) nor did it influence the ability of HCE to cap G21 transcripts associated with high salt washed complexes (Figure 34). Regardless of the mechanism of cap inhibition by H$_2$O$_2$, these results clearly indicate that co-transcriptional capping is not required for pausing. However, as all capped RNAs were associated with complexes which paused normally, it is possible that capping and pausing share a Cdk7-dependent requirement.

To determine if DSIF, NELF, or Gdown1 were involved in the effect of THZ1 on pausing, I added DSIF and NELF, NELF alone, or Gdown1 to control or THZ1-treated complexes after a high or low salt wash. Flavopiridol was used in all reactions to block the transition into productive elongation. After 5 minutes of add-back incubation, elongation was continued for 3 minutes to examine changes in polymerase pausing. Regardless of THZ1 treatment, DSIF and NELF, but not NELF alone, slowed elongation of high salt washed complexes while Gdown1 had no significant effect (Figure 35A).
Figure 35. Pausing factor add-backs to high or low salt washed complexes ± THZ1
(A) Elongation complexes were generated by preincubation with flavopiridol alone (Control) or in combination with THZ1 and a 30 s limiting UTP/CTP pulse. Complexes were isolated by high or low salt wash, incubated for 5 min with mock (-), DSIF and NELF (DN), NELF only (N), or Gdown1 (G) add-backs, and chased for 3 min. Final concentrations during the chase were 1 μM for both flavopiridol and THZ1. 6% Urea-PAGE. (B) Profiles from the bracketed region (20-550 nt, left to right) of control (black) or THZ1 (red) low salt wash mock add-backs. The vertical axis represents relative phosphorimage signal.
Figure 36. Profiles of pausing factor add-backs to low salt washed complexes ± THZ1
Profiles from the bracketed region (20-550 nt, left to right) of Figure 35A of pausing factor add-backs to low salt washed elongation complexes generated in the presence of flavopiridol alone (A, C, E) or in combination with THZ1 (B, D, F). Mock (black) add-backs are compared with DSIF/NELF (orange) (A-B), NELF only (pink) (C-D), or Gdown1 (green) (E-F) add-backs. The vertical axes represent relative phosphorimage signal intensity and all plots are the same scale.
expected, low salt washed complexes elongated more efficiently due to retention of Pol II-associated factors. While THZ1 had no effect on high salt washed complexes, THZ1-treated low salt washed complexes generated longer transcripts and more complexes reached run-off at 508 nt than control complexes. Inhibition of pausing in these THZ1 complexes was clearly seen in lane profiles (Figure 35B) and in agreement with previous results (Nilson et al., 2015).

Whole gel profiles were also made for low salt wash add-back lanes to enable side-by-side comparison of changes in elongation between control and THZ1 complexes (Figure 36). Preincubation of DSIF and NELF with low salt washed control complexes greatly shortened the transcripts generated (Figure 36A). Notably, preincubation with NELF alone caused a similar increase in pausing (Figure 36C). This suggests that DSIF, but not NELF, survives the low salt wash. When preincubated with low salt washed THZ1 complexes, the pausing effects of DSIF and NELF (Figure 36B) and especially NELF alone (Figure 36D) were significantly reduced. This suggests that DSIF was not properly loaded before the wash and that elongation complexes generated in the presence of THZ1 are highly resistant to addition of DSIF and NELF. Gdown1 had an expected negative effect on low salt washed control complexes due to the presence of GNAF (Cheng et al., 2012) and this effect was seen regardless of THZ1 (Figure 36E-F). Because the THZ1 effect on DSIF loading is salt-sensitive, these results suggest that THZ1 causes inappropriate retention of a blocking factor. This factor could also block capping.

Effects of THZ1 seen in vitro also occur in cells

To determine if the effects of THZ1 in vitro on capping and pausing also occur in cells, I performed nuclear run-on experiments and measured the lengths of transcripts associated with polymerases engaged with chromatin. I isolated nuclei from adherent HeLa S3 cells that were treated with DMSO, 1 μM flavopiridol or 1 μM THZ1 for 1 hour. One hour of 1 μM THZ1 significantly reduced Pol II CTD phosphorylation both in vitro and in cells (Nilson et al., 2015). Elongation of engaged polymerases was carried out with 500 μM cold ATP, UTP, and GTP, 2 μM cold CTP, and 0.33 μM α-³²P-CTP, conditions which resemble a limiting-CTP pulse in vitro. Sarkosyl was used to prevent new initiation and to eliminate effects of negative factors and histones during the short run-on reactions so that paused Pol II could be detected. Additionally, α-amanitin was
Cap status determination of nascent transcripts generated by nuclear run-ons performed for 4 min with 500 μM ATP/UTP/GTP, 2 μM cold CTP, and 0.33 μM α-³²P-CTP in the absence or presence of 4 μg/ml α-amanitin using nuclei from adherent HeLa S3 cells treated 1 h with DMSO only (black), 1 μM flavopiridol (blue), or 1 μM THZ1 (red). T: total transcripts. C: capped transcripts recovered with anti-2,2,7-trimethylguanosine beads. 6% Urea-PAGE.

added to every other reaction to specifically inhibit Pol II, but not Pol I or III. The pattern of total transcripts from control cells clearly showed Pol II (amanitin-sensitive) transcripts of the appropriate size to have been generated by promoter proximal paused polymerases (Figure 37). After cap methylation, a fraction of the transcripts could be recovered by anti-2,2,7-trimethylguanosine beads. In this experiment, the efficiency of binding to the beads was not determined so quantification of the percent of transcripts capped was not possible. As expected, no capped nascent transcripts were detected when Pol II was inhibited.
I generated profiles of the resulting Pol II-dependent total or capped nascent transcripts by first normalizing each sample pair to cold nuclear RNAs visible in the ethidium bromide stained gel to account for variations in loading and then taking the difference between run-ons performed with or without α-amanitin (Figure 38). These profiles were carefully aligned using the discreet bands of α-amanitin-insensitive transcripts. Profile analysis of the Pol II transcripts showed the expected effect of flavopiridol: pausing was intact and transcripts were capped (Figure 38C). THZ1 caused a reduction in pausing, a slight increase in longer transcripts compared to flavopiridol, and significantly reduced capping efficiency (Figure 38A, D). These results strongly indicate that THZ1 functions in a similar manner in vitro and in cells.

After performing the above experiments, I discovered that nuclear run-on resolution could be dramatically improved by adding only α-³²P-CTP to isolated nuclei (Figure 39). I performed these improved nuclear walk-ons on nuclei isolated from...
Figure 39. THZ1 inhibits proper mRNA capping and Pol II elongation in cells

(A) Cap status determination of transcripts generated by 30 s pulse with limiting α-32P-CTP or by nuclear walk-ons performed in the absence or presence of 4 μg/ml α-amanitin using nuclei from suspension HeLa cells treated 1 h with DMSO only (black), 1 μM flavopiridol (blue), or 1 μM THZ1 (red). snRNAs with (U2, U1, U4, U5) and without (5.8S, U6) trimethylguanosine cap structures are indicated. 9% Urea-PAGE.

(B) Profiles of relative phosphorimage signal from the bracketed region. Total profiles were generated by combining signals from capped and uncapped lanes. All plots have the same vertical scale.
suspension HeLa cells treated with DMSO, 1 μM flavopiridol or 1 μM THZ1 for 1 hour. In this experiment, the efficiency of binding to the beads was nearly quantitative as demonstrated by phosphorimaging of control pulse lanes and recovery of Pol II-transcribed small nuclear RNAs (snRNAs) with appropriate cap structures (Figure 39A). Amanitin-sensitive Pol II transcript profiles were generated by taking differences of lanes with and without α-amanitin; total Pol II profiles were made by adding capped and uncapped Pol II profiles (Figure 39B). Control nuclei had transcripts of the appropriate size (30-50 nt) for promoter proximal pausing and a large fraction of these were capped. Flavopiridol treatment resulted in increases to both the amount of pausing and the fraction of capped transcripts. An apparent reduction of short transcripts was seen after THZ1 treatment, indicating a reduction of promoter proximal pausing, and the remaining transcripts were mostly uncapped. An additional RNA population was seen around 20 nt, but these RNAs, once isolated, were unresponsive to calf intestinal alkaline phosphatase or high amounts of HCE in a separate experiment (Figure 26), suggesting that they were the result of 5′ degradation to the edge of the polymerase. Overall, I conclude that THZ1 leads to reduced pausing and capping both in vitro and in cells.

Discussion

The Cdk7 inhibitor THZ1 has opened a new window through which to observe mRNA capping and the early events of Pol II elongation control. My findings help to further define the timing of enzymatic activities and the required exchange of factors that results in capping and promoter proximal pausing (Figure 40). I propose that Cdk7 is essential for the timely execution of these transitions. In this model, CTD phosphorylation occurs swiftly and is followed by capping, pausing by DSIF and NELF, and the P-TEFb-dependent transition into productive elongation. After THZ1 treatment, CTD phosphorylation does not occur and early capping and loading of DSIF and NELF is defective. Lack of CTD phosphorylation did not affect pausing by DSIF and NELF and had only a small effect on capping using recombinant proteins on high salt washed complexes. Therefore, the major effects seen must be caused by other factors present in nuclear extract. The loss of DSIF is sufficient to inhibit productive elongation because DSIF is the functional target of P-TEFb. The sum of these disruptions in an
My results reveal that cap addition appears to occur preferentially at a “sweet spot” near the RNA exit pore and access of the capping enzyme is modulated by a Cdk7-dependent factor. While I do not know the identity of this factor, it is possible that it also influences pausing. These findings disagree with a proposed checkpoint model developed primarily in fission yeast, in which the Spt5 subunit of DSIF both stimulates and provides a temporal window for capping (Pei et al., 2003). While DSIF moderately stimulates nascent transcript capping by mammalian capping enzyme in vitro (Mandal...
et al., 2004) and its loading onto elongation complexes could be a shared THZ1-sensitive requirement for capping and pausing, DSIF does not efficiently associate with elongation complexes until after nascent transcripts have extended significantly beyond the ~21 nt sweet spot (Cheng and Price, 2008). Additionally, Spt5 associates with chromatin downstream of capping enzymes genome-wide in budding yeast (Lidschreiber et al., 2013). While these results suggest that sweet spot guanylylation and DSIF are unlikely to be co-dependent in mammals, the delayed, inefficient capping of longer nascent transcripts could rely on the stimulatory effects of Pol II and Spt5 CTD phosphorylation and such capping might share more of its mechanisms with yeast.

After THZ1 treatment, I observed salt-sensitive association of unknown factors which inhibited capping and DSIF loading. These could be initiation factors that are inappropriately retained by the transcription complex in the absence of CTD phosphorylation. One plausible candidate for the pausing inhibitor is TFIIE as its archaeal homolog TFE competes with DSIF for binding to archaeal Pol II (Grohmann et al., 2011; Martinez-Rucobo et al., 2011) and its retention increases after Cdk7 inhibition in human cells (Kelso et al., 2014; Larochelle et al., 2012). TFIIE is also a direct target for Cdk7 phosphorylation in vitro (Larochelle et al., 2012). Multiple lines of evidence suggest that Mediator could be involved. In mammals, the interaction of TFIIE and TFIIH with the preinitiation complex is stabilized by Mediator (Poss et al., 2013). A very recent structural study demonstrated that the yeast capping enzymes interact directly with the RNA exit channel and Rpb4/7 (Martinez-Rucobo et al., 2015) and this region of Pol II is normally occupied by Mediator during initiation (Plaschka et al., 2015).

Importantly, the complex interaction of Mediator with Pol II is regulated in part by CTD phosphorylation (Jeronimo and Robert, 2014; Wong et al., 2014). Occam’s razor would suggest that the inhibitors of pausing and capping are the same factor. Additionally, this could explain why CTD phosphorylation has only a minor direct effect on capping of transcripts in isolated elongation complexes but a larger effect after initiation in extract with Mediator (Nilson et al., 2015). Overall, I believe an inhibitor of capping and pausing, potentially Mediator, is present during initiation and normally requires Cdk7 activity to dissociate near G21. This departure could then enable DSIF loading which
would allow NELF-mediated pausing, cap addition as transcripts emerge from the Pol II exit channel, and ultimately productive elongation.

These results help explain the characteristic effects of THZ1 seen in cells and this information highlights how THZ1 might be useful in cancer treatment. The devastating effects of THZ1 on transcription seen in vitro would be partially mitigated by using lower concentrations of the compound. Of 1,151 cancer cell lines tested, THZ1 exhibited antiproliferative effects on more than half at concentrations below 200 nM (Kwiatkowski et al., 2014). Using a sub-saturating amount of the covalent inhibitor would ensure that some active Cdk7 would remain in each cell to drive sufficient transcription. An effective treatment would have to balance treatment time, the concentration of the compound, and the relative amount of compound compared to its target. IC50s alone cannot be used as while initial non-covalent binding is concentration-dependent, covalent bond formation is slow and its success depends primarily on the off rate of the compound before covalent linkage. Notably, an analog of THZ1 has already been developed with improved pharmacokinetic features and a 5-fold improved half-life in vivo (Wang et al., 2015).

THZ1 appears to achieve its effect on cancer cells by reducing expression of transcription factors required for cell proliferation, such as RUNX1 in T-cell acute lymphoblastic leukemia (Kwiatkowski et al., 2014), MYCN in neuroblastoma (Chipumuro et al., 2014), MYC family members and neuroendocrine lineage-specific factors in small cell lung cancer (Christensen et al., 2014), and FOSL and SOX9 in triple-negative breast cancer (Wang et al., 2015). The common thread is that all target genes are driven by super-enhancers that normally function as amplifiers of cell-specific transcription factors in rapidly growing cells. Super-enhancers are built upon regions of the genome containing paused Pol II (Arner et al., 2015; Core et al., 2014) and I suggest that THZ1-induced pause defects lead to reduced occupancy of Pol II and the concomitant reduction in transcription factors in these enhancers. If super-enhancers amplify gene expression through the multiplicative effect of a large number of occupied sub-enhancers, impairing each sub-enhancer slightly could have a significantly larger overall effect on target gene expression. In this way, THZ1 could selectively target the super-enhancer-driven transcriptional addiction seen in cancer.
CHAPTER 4: TRANSCRIPTIONAL RESPONSES TO OXIDATIVE STRESS

This chapter focuses on the rapid and widespread effects of hydrogen peroxide on transcription.

Introduction

Over time, organisms have evolved mechanisms to adapt to a variety of stresses. Cells must be able to rapidly transcribe messages in response to insults such as heat shock, hypoxia, UV radiation, and oxidative stress. Hydrogen peroxide ($H_2O_2$) can be formed exogenously from the interactions of water with pollutants or ionizing radiation, or endogenously by dismutation of superoxide, a natural byproduct of aerobic respiration (Veal et al., 2007). Hydroxyl radicals from $H_2O_2$ can attack both the deoxyribose backbone and bases of DNA and the resulting adducts are the most abundant form of DNA damage in cells (Beckman and Ames, 1997). If not corrected by nucleotide excision repair (NER) machinery, oxidative lesions can lead to miscoding and DNA polymerase arrest during replication, or RNA polymerase arrest during transcription (Laine and Egly, 2006). In addition to reacting with DNA, hydrogen peroxide can oxidize sulfur-containing groups in cysteine (Rudyk and Eaton, 2014) and methionine residues (Ghesquiere et al., 2011). Complex networks exist to sense and respond to varying concentrations of $H_2O_2$ (Veal et al., 2007).

The main peroxide clearance enzymes in eukaryotes are catalase, glutathione peroxidases, and thioredoxin peroxidases (also known as peroxiredoxins) (Veal et al., 2007). In yeast, these proteins are regulated primarily by controlling their expression through a complex network of transcription factors and chromatin remodelers (Garcia et al., 2016). In higher eukaryotes, peroxide-targeting enzymes are also regulated by post-translational modifications in a peroxide concentration-dependent manner. Low $H_2O_2$ concentrations induce tyrosine kinases which phosphorylate catalase, increasing its activity (Patterson et al., 2015). Catalase is ubiquitinated and degraded in response to irrecoverably high levels of peroxide, however, as part of a coordinated cell-death response (Veal et al., 2007). Excessive levels of hydrogen peroxide also lead to superoxidation and inactivation of peroxiredoxins although this is offset by sestrin and sulfiredoxin proteins, whose expression is induced by oxidative stress (Veal et al., 2007). Ultimately, this combination of peroxide clearance proteins and their
concentration-dependent activities results in a system where H\textsubscript{2}O\textsubscript{2} up to a certain level is successfully removed and cell death is triggered when this threshold is breached (Sena and Chandel, 2012).

In the event of DNA damage after oxidative stress, cells rapidly downregulate new initiation and activate transcription-coupled repair pathways to rescue RNA polymerases that stall at lesions (Laine and Egly, 2006). Nuclear extracts from HeLa S3 cells treated in media with 0.3 mM H\textsubscript{2}O\textsubscript{2} for 15 minutes exhibited no in vitro transcriptional activity and only when these cells were allowed to recover for 30 minutes were their extracts able to synthesize new RNA (Heine et al., 2008). Ubiquitination of Pol II subunit Rpb1 and hyperphosphorylation of its CTD Ser5 residues correlated with this peroxide-induced loss of initiation activity, which was not alleviated by proteasome or CTD kinase inhibition (Heine et al., 2008). Notably, methionines within Rpb1 are also directly oxidized by H\textsubscript{2}O\textsubscript{2} (Ghesquiere et al., 2011). Even when Pol II is undamaged, it cannot traverse DNA templates which contain 5-hydroxyuracil, 7,8-dihydro-8-oxoguanine, or thymine glycol without assistance from factors present in nuclear extracts, and addback of purified TFIIF, Elongin, or the transcription-coupled NER protein CSB prevented many of these polymerases from stalling, albeit with some NTP misincorporation (Charlet-Berguerand et al., 2006). Despite the peroxide-mediated block to initiation, treatment of fibroblasts with 250 \text gamma M H\textsubscript{2}O\textsubscript{2} has been reported to induce about one hundred genes, many of which are involved in DNA repair and stress response (Kyng et al., 2003). Importantly, cells lacking CSB had this transcriptional response delayed from 15 min to 6 h after H\textsubscript{2}O\textsubscript{2} treatment (Kyng et al., 2003).

A recent study of transcription in MRC5 cells demonstrated that hydrogen peroxide induces genome-wide changes in RNA synthesis and Pol II ChIP-Seq occupancy (Giannakakis et al., 2015). The global changes observed in this study conflict with a prevailing model where oxidative stress induces targeted and controlled changes in transcription (Sena and Chandel, 2012; Veal et al., 2007) through transcription factor signaling pathways (Marinho et al., 2014; Patterson et al., 2015) and histone modification (Chervona and Costa, 2012; Sanders et al., 2013). Because my earlier research demonstrated that H\textsubscript{2}O\textsubscript{2} could inhibit capping and initiation, but not
pausing in transcription reactions performed with HeLa nuclear extracts (Nilson et al., 2015), I sought to characterize this rapid response using nuclear walk-on assays and ChIP-Seq. I discovered that hydrogen peroxide treatment leads to a massive increase in transcriptionally engaged Pol II near promoters after as little as two minutes and these polymerases act differently in the presence or absence of P-TEFb activity. Additionally, P-TEFb governed the activities of Pol II at enhancers as well as promoters and appeared to influence elongation through the -1 nucleosome in H₂O₂-treated cells. Not only do these results help explain how cells respond to oxidative stress, this research highlights the importance of elongation control machinery in bidirectional transcription.

**Materials and Methods**

*In vitro transcription*

The use of CMV templates and HeLa nuclear extract (HNE) to study transcription in vitro was described in Chapters 2 and 3. Template DNA from CMV -800 bp to +508 bp was incubated for 30 min with 1 μl/rxn HNE in the presence of 60 mM KCl, 5 mM MgCl₂, 20 mM HEPES pH 7.6, 1 mM DTT, and 0.5 U/μl SUPERase-In. Initiation was accomplished with a 30 s pulse containing 60 mM KCl, 5 mM MgCl₂, 20 mM HEPES pH 7.6, 0.21 μM α⁻³²P-CTP, and 500 μM ATP/UTP/GTP (limiting CTP). Elongation complexes were either stopped by addition of EDTA to 20 mM or chased for 3 min with 500 μM CTP. Labeled transcripts were extracted with phenol, precipitated, and analyzed as described in previous chapters.

*Nuclear walk-on*

HeLa cells were grown at 37°C and 5% CO₂ to 80% confluency in T150 flasks in 30 ml DMEM supplemented with 10% FBS. One hour before H₂O₂ treatment, 10 ml of media was removed, spiked with 1:1,000 final volume of DMSO containing 1 mM of flavopiridol as indicated (1 μM final), and re-added to cells. 10.3 M H₂O₂ (Fisher H325) was diluted in H₂O immediately before use and added to cells for times and final concentrations as indicated. All treatment times were staggered to ensure harvest of each sample was separated by 1-2 min. All subsequent cell handling was performed on wet ice and all buffers were ice cold. After compound incubation, each flask was removed from the incubator and over the next 20 s, media was dumped into a waste container, 50 ml of
PBS was added and quickly mixed, PBS was dumped out, and 15 ml lysis buffer (10 mM Tris pH 7.6, 320 mM sucrose, 0.5% Triton X-100, 1 mM spermine (Sigma S1141), 1 mM spermidine (Sigma S2501), 1 mM EDTA, 1 mM DTT, 0.004 U/μl SUPERase-In, and 0.1% isopropanol-saturated PMSF) was added. Cells were incubated for 10-15 min on ice and lysis was monitored by phase-contrast microscopy. Lysed cells were scraped, pelleted at 1200 × g for 5 min, and thoroughly resuspended in a mixture of 1 ml lysis buffer and 2 ml sucrose cushion (10 mM Tris pH 7.6, 1.9 M sucrose, 1 mM spermine, 1 mM spermidine, 0.1 mM EDTA, 1 mM DTT, 0.004 U/μl SUPERase-In, and 0.1% isopropanol-saturated PMSF). This homogenate was carefully layered over a 2 ml sucrose cushion and spun at 30,000 × g for 45 min. Nuclei were thoroughly resuspended in 100 μl storage buffer per T150 flask, homogenized using a loose pestle and a 1 ml Dounce, aliquoted, and frozen at -80°.

For each reaction, 1 × 10⁵ nuclei were diluted to 20 μl in storage buffer lacking glycerol and incubated for 3 min at 30° with an equal volume of NRO buffer (10 mM Tris pH 7.6, 1% sarkosyl, 300 mM KC₃H₃O₃, 5 mM Mg(C₂H₃O₂)₂, 5 mM DTT, 0.5 U/μl SUPERase-In, and 0.33 μM α-³²P-CTP). Reactions were stopped by addition of EDTA to 20 mM and labeled transcripts were extracted with Trizol LS and precipitated with 95% ethanol and 500 mM NH₄C₂H₃O₂. Nascent transcripts were either analyzed immediately or resuspended in a buffer containing 20 mM Tris pH 7.6 and 0.5 U/μl SUPERase-In and separated by cap status as described in Chapter 2.

**ChIP-Seq**

For ChIP-Seq in adherent HeLa cells, HeLa cells were grown at 37° and 5% CO₂ to 80% confluency in T150 flasks in 30 ml DMEM supplemented with 10% FBS. One hour before H₂O₂ treatment, 10 ml of media was removed, spiked with 1:1,000 final volume of DMSO containing 1 mM of flavopiridol as indicated (1 μM final), and re-added to cells. 10.3 M H₂O₂ was diluted in H₂O immediately before use and added to cells (0.3 mM final) for 3, 10, 30, and 100 min. All treatment times were staggered to ensure crosslinking of each sample was separated by 1 min. Cells were removed from the incubator and crosslinked promptly by adding 16% paraformaldehyde to a final concentration of 1% in media. Crosslinking was allowed to proceed for 10 min before Tris pH 7.6 addition to 1.33 M. All solutions used for steps after crosslinking and before elution were ice cold.
and supplemented with EDTA-free protease inhibitor cocktail and 0.1% isopropanol-saturated PMSF. Cells were pelleted at 1000 × g for 5 min at 4° and rinsed twice with PBS. Decanted, rinsed pellets were stored at -80°.

One T150 flask (~1.5 × 10⁷ cells) was used per ChIP. Pellets were thawed in 1 ml 0.2% sarkosyl buffer, homogenized with 10 loose pestle strokes in a 1 ml Dounce, and sonicated in 14 ml polystyrene tubes on wet ice with a Fisher Model 550 Sonic Dismembrator for 15 cycles of 30 s at intensity 3.2-4, followed by 60 s off. After sonication, 100 μl 10% Triton X-100 was added and samples were pelleted at 16,000 × g for 15 min at 4°. Supernatants were stored at -80°. For each ChIP, 100 μl M-280 Sheep anti-Rabbit IgG Dynabeads (Novex 11203D) were washed with 100 μl PBS, blocked in blocked in 100 μl PBS containing 1 mg/ml BSA for 1 h at 4° with rotation, washed again with 100 μl PBS, and incubated with 10 μg Pol II antibody for 2 h at 4° with rotation. Antibody-beads were washed twice with 200 μl wash buffer A, then resuspended with sonicated supernatants and incubated overnight at 4° with rotation. 1% of each sample was set aside for input sequencing. Samples were washed once with 200 μl wash buffer A, 200 μl wash buffer B, and 200 μl wash buffer C, and twice with 200 μl rinse buffer. Samples were incubated with 200 μl elution buffer for 2 h at 65° to reverse crosslinks. Eluates were treated with 2 μl RNase A for 30 min at 37°, then 2 μl Proteinase K for 30 min at 50°. DNA was isolated by MinElute PCR Purification Kit and libraries were prepared using a NEXTflex Rapid DNA-Seq Kit. Sequencing was performed by the Iowa Institute of Human Genetics Genomics Division on an Illumina HiSeq 2500 using 125 bp paired-end reads.

Raw paired sequences were first aligned to the UCSC hg19 assembly using Bowtie 2.2.6 (--no-mixed --no-discardant --reorder) (Langmead and Salzberg, 2012). Human-mapped reads were compiled using MACS 2.0.10 (--format BAMPE --bdg) (Zhang et al., 2008), compressed (Kent et al., 2010), and visualized on the UCSC Genome Browser (Kent et al., 2002). ChIP-Seq tracks were normalized by dividing the number of hg19-mapped reads within each sample by the average number of hg19-mapped reads across samples within an experiment. Additional normalization was skipped for metagene and heatmap analyses, which were otherwise performed as in Chapter 2.
Results

Hydrogen peroxide inhibits capping, productive elongation, and initiation in vitro

Previously performed experiments suggested that hydrogen peroxide could influence both transcription and mRNA processing. Unpublished work performed in the Price Lab by Todd Adamson showed that the transition into productive elongation was redox-sensitive in vitro: DTT increased the number of elongation complexes that reached run-off while H₂O₂ reduced them (data not shown). Follow up experiments showed that at a narrow concentration range of H₂O₂ could inhibit capping without affecting promoter-proximal pausing, and this cap inhibition was unlikely to be from direct impairment of the RNA triphosphatase or guanylyltransferase activities of HCE (Figures 33-34) (Nilson et al., 2015).

To gain a better understanding of the effects of H₂O₂ on transcription, increasing amounts of peroxide were added to in vitro transcription reactions and changes were observed in initiation and early elongation (pulse), and in productive elongation (pulse, then chase) (Figure 41). Preinitiation complexes (PICs) were formed with a 30 minute preincubation that included HeLa nuclear extract, template DNA, and a range of 0.02 to 200 mM H₂O₂. Initiation was accomplished with a 30 second pulse containing 500 μM ATP/UTP/GTP and a limiting amount of α-³²P-CTP. In the absence of H₂O₂, a typical CTP-limiting pattern of transcripts was seen; elongation complexes stopped after 16, 21, 25, 31, 49, 54, and 56 nt and capping was observed on transcripts greater than 20 nt in length. Peroxide inhibited co-transcriptional capping partially at 0.2 mM and completely at 2 mM. Higher peroxide concentrations also reduced the amounts of transcripts generated slightly at 2 mM and significantly at 20 mM. No incorporation, and thus no initiation was detected after 200 mM H₂O₂.

To evaluate the effects of peroxide on productive elongation, pulsed complexes were chased for 3 minutes by raising CTP levels to 500 μM. Because all components of HeLa nuclear extract were present (including DSIF and P-TEFb), elongation complexes could efficiently reach run-off at 508 nt in the absence of H₂O₂. As evidenced by decreasing run-off transcripts, productive elongation was inhibited by peroxide slightly at 0.2 mM, significantly at 2 mM, and completely at 20 and 200 mM. While run-off transcripts were reduced at 0.2 and 2 mM H₂O₂, the signal from slower elongation
Figure 41. Effects of H$_2$O$_2$ on transcription in vitro

HeLa nuclear extract, template DNA, and H$_2$O$_2$ as indicated were preincubated for 30 min, pulsed for 30 s with limiting $\alpha$-$^{32}$P-CTP, and chased as indicated for 3 min with 500 $\mu$M CTP. RO: run-off. Lengths of limiting-CTP transcripts are indicated; brackets indicate sizes of capped and uncapped transcripts. In the presence of nuclear extract, tRNAs incorporate $\alpha$-$^{32}$P-CTP independently of Pol II transcription. Experiment designed and supervised by Kyle Nilson, and performed by Nicholas Mullen. 6% Urea-PAGE.
complexes was not apparently reduced. These result suggests that moderate peroxide concentrations that impair co-transcriptional capping also inhibit Pol II elongation in vitro. It remains unclear which elongation factors contribute to these observed effects.

$H_2O_2$ induces a global transcriptional response in cells

To measure the effects of peroxide on Pol II transcription in cells, nuclear walk-ons were performed in the presence and absence of $\alpha$-amanitin, a potent Pol II-specific inhibitor, using adherent HeLa cells treated with $H_2O_2$ over a range of times and concentrations (Figure 42). A number of optimizations were applied to the nuclei isolation protocol (Figure 42A). By switching to adherent cells, I could dump off the media, rinse the cells with ice-cold PBS, and pour on lysis buffer in about 20 seconds. This is in stark contrast to suspension cells, which must be pelleted for about 5 minutes, carefully decanted, rinsed with PBS, and spun again before lysis buffer can be added, an elaborate dance that takes nearly 15 minutes. To eliminate any possibility of unwanted Pol II elongation during the isolation, 1 mM EDTA was also added to the lysis buffer. Because this lysis buffer contains 0.5% Triton X-100, EDTA readily entered the nucleus and rapidly inactivated any magnesium-dependent proteins. To stabilize chromatin in the absence of magnesium, 1 mM of spermine and spermidine were included. After lysis, cellular homogenates were adjusted to about 1.4 M sucrose, carefully layered over a 1.9 M sucrose cushion, and centrifuged at 30,000 × g for 45 minutes. At this concentration of sucrose, nuclei were efficiently separated from other organelles. Because most endogenous NTPs diffuse out of the nucleus in the presence of detergent, they too were removed by the sucrose pad. After centrifugation, sucrose layers were aspirated in two steps to minimize contamination. Finally, nuclei were resuspended in a storage buffer which contained magnesium to enable elongation, and 25% glycerol to prevent damage during freezing. While each experimental condition starts with flasks containing equivalent numbers of cells, variation can be introduced by incomplete removal of the viscous sucrose cushion and by incomplete resuspension in the viscous storage buffer. This is controlled by quantifying nuclear RNAs visualized by ethidium bromide and using these values to normalize the phosphorimage profiles of walk-on nascent transcripts. Phosphorimage profiles of nascent transcripts are then aligned using the discreet bands of $\alpha$-amanitin-insensitive transcripts. Finally, Pol II signal is
Figure 42. Effects of \( \text{H}_2\text{O}_2 \) on global transcription detected by nuclear walk-on

(A) Experimental outline of the nuclear walk-on assay. (B) Nuclear walk-ons with nuclei isolated from adherent HeLa cells treated with \( \text{H}_2\text{O}_2 \) for the times indicated. Every other reaction contains 2 \( \mu \text{g/ml} \) \( \alpha \)-amanitin. 6% Urea-PAGE. (C-D) Whole-gel profiles of Pol II nascent transcripts. First, nascent transcripts were normalized using nuclear RNAs. Then, \( \alpha \)-amanitin-insensitive transcripts were subtracted from total transcripts. The vertical axes represent relative phosphorimage signal intensity and both plots are the same scale.
calculated by subtracting each α-amanitin profile from its paired total transcript profile. Importantly, the distributions and intensities of these signals are highly reproducible and can be used to quantitatively assess Pol II engagement.

As expected, walk-ons performed using untreated nuclei generated a population of amanitin-sensitive 30-50 nt transcripts and these correspond with promoter-proximal paused Pol II (Figure 42B). Additionally, there was little accumulation of ~20 nt labeled RNAs from degraded transcripts still associated with elongation-competent polymerases. In addition to the 30-50 nt transcripts, a subpopulation of slightly longer transcripts were detected in the cells treated 10 minutes with 0.3 mM H₂O₂ (Figure 42C). After 30 minutes, a drastic increase was observed in short transcript signal and this peak was shifted down the gel. This result could indicate that Pol II is pausing more upstream than normal. Alternatively, backtracking and transcript cleavage by TFIIS could result in this shift and still produce an elongation-competent polymerase. Longer transcript signals also increased slightly. After 100 minutes, the signals in the 30-50 nt region continued to increase while longer transcripts began to decline in levels. Interestingly, the intensity, but not the pattern of transcripts changed after 30 minutes of 1 mM H₂O₂; almost no response was observed at 0.1 mM H₂O₂ (Figure 42D). While the positions of Pol I and III do not appear to change after peroxide treatment, as evidenced by the lack of migration of the bands in α-amanitin-containing lanes, their activities appear to be inhibited by H₂O₂ in a concentration-dependent manner (Figure 42B).

Two overall effects were seen after peroxide: the brief appearance of slowly elongating Pol II complexes after 10 minutes, and the increase in promoter-proximal, but not necessarily paused Pol II after 30 and 100 minutes. The degree of this transcriptional response appears to be modulated by H₂O₂ concentration but this response, once initiated, takes a consistent amount of time to execute. The minimal change observed after 30 minutes of 0.1 mM H₂O₂ could be due to successful clearance of peroxide by catalase and other enzymes, alleviating the need for a global response. When observed by confocal microscopy, adherent HeLa cells continued growing and remained attached with 0.3 mM or lower concentrations of H₂O₂ over time; 1 mM H₂O₂, however, triggered widespread cell death sometime between 2.5 and 18 hours.
Figure 43. Effects of H$_2$O$_2$ in the absence and presence of flavopiridol

(A) Nuclear walk-ons with nuclei isolated from adherent HeLa cells treated with or without 1 μM flavopiridol for 1 h, then with 0.3 mM H$_2$O$_2$ for the times indicated. 6% Urea-PAGE. (B) Whole-gel profiles of total nascent transcripts normalized using nuclear RNAs. The vertical axes represent relative phosphorimage signal intensity and both plots are scaled to have comparable no H$_2$O$_2$ (black) 30-50 nt peak heights.
P-TEF\textsubscript{b} contributes to a rapid response to H\textsubscript{2}O\textsubscript{2}

I next performed nuclear walk-ons using nuclei from HeLa cells treated with 0.3 mM H\textsubscript{2}O\textsubscript{2} for 2, 5, 10, 20, or 50 minutes. Surprisingly, the signal from short nascent transcripts more than doubled after only 2 minutes of peroxide treatment (Figure 43A). These transcripts were also slightly longer than those generated by promoter-proximal paused Pol II in untreated cells. After 5 minutes, the amount of radiolabeled transcripts below 150 nt in length was comparable to untreated levels as shown by the corrected gel profiles (Figure 43B). These remaining transcripts were also generally longer than those seen after 2 minutes of peroxide, and this apparent elongation continued over the 10 and 20 minute time points. A new population of 30-50 nt transcripts emerged after 20 minutes and remained after 50 minutes; as observed previously, these transcripts were slightly shorter than those associated with pausing in untreated cells. After 50 minutes, transcripts from potentially elongating polymerases began to disappear. These results indicate that the transcriptional response to hydrogen peroxide is extremely rapid.

To determine if P-TEF\textsubscript{b} contributes to this rapid response, I repeated the above experiment using cells treated with 1 \muM flavopiridol for 1 hour prior to addition of hydrogen peroxide. As would be expected after treatment with a potent P-TEF\textsubscript{b} inhibitor, flavopiridol led to an increase in promoter-proximal pausing in cells. Unexpectedly, flavopiridol accentuated the increase in short transcripts seen after 2 minutes of 0.3 mM H\textsubscript{2}O\textsubscript{2} (Figure 43A). The magnitude of this increase between no H\textsubscript{2}O\textsubscript{2} and 2 minutes of H\textsubscript{2}O\textsubscript{2}, however, was similar in cells treated with or without flavopiridol (Figure 43B). After 5 minutes of peroxide, the entire peak of transcripts shifted up the gel by about 10 nt but did not decrease in intensity, as was observed in control cells at this time point. These slowly growing transcripts continued to lengthen over the next two time points. Transcripts between 30-50 nt in length began to reappear after 20 minutes and after 50 minutes of H\textsubscript{2}O\textsubscript{2}, these short RNAs were the predominant form present. In summary, flavopiridol prevented a disappearance of transcripts between 2 and 5 minutes of peroxide, but did not influence the restoration of 30-50 nt transcripts.
Hydrogen peroxide caused the number of promoter-proximal polymerases to more than double after 2 minutes, which could be due to increased initiation or from inhibition of termination, both of which would allow more polymerases to accumulate near promoters. Because the magnitude of this increase was essentially unchanged by flavopiridol, I conclude that inhibition of termination is more likely. If the rate of termination is the same before H₂O₂ in both control and flavopiridol-treated cells, a uniform inhibition of termination would cause the observed proportional change. If initiation was stimulated after H₂O₂, this theoretical increase in initiation would have to be stronger in flavopiridol-treated cells to produce the observed result. Because the disappearance of transcripts between 2 and 5 minutes of peroxide treatment occurred only in control cells and not flavopiridol-treated cells, I conclude that the majority of polymerases that accumulate after H₂O₂ are acted upon by P-TEFb and potentially enter productive elongation. This result would not be obtained if DNA damage was responsible for the accumulation of arrested polymerases. The few polymerases which remain after 5 minutes of peroxide treatment in control cells appear to slowly elongate at the same rate as the polymerases in flavopiridol-treated cells, and this rate is far below the ~3 kb/min expected from productively elongating Pol II. These non-productively elongating complexes could have associated pausing factors.

**Nascent transcript cap status after H₂O₂ treatment**

Because hydrogen peroxide inhibited co-transcriptional capping in vitro, I suspected that capping would be inhibited in cells. To test this, I repeated the previous nuclear walk-on experiment and evaluated the cap status of the resulting nascent transcripts. After Trizol extraction, guanylylated transcripts were methylated by recombinant human cap methyltransferase and isolated using an anti-2,2,7-trimethylguanosine antibody that recognizes the mature m⁷G cap. While cap binding was not quantitative—capped snRNAs were not recovered with 100% efficiency as seen by ethidium bromide—binding was consistent across the experiments (Figure 44). Because of this, and because control and flavopiridol experiments were on separate gels, normalized phosphorimage profiles of nascent transcripts were scaled to have similar 30-50 nt peak heights in the absence of H₂O₂ (Figure 45).
Figure 44. Cap status of nascent transcripts after \( \text{H}_2\text{O}_2 \) treatment

Nuclear walk-ons with nuclei isolated from adherent HeLa cells treated with or without 1 \( \mu \text{M} \) flavopiridol for 1 h, then with 0.3 mM \( \text{H}_2\text{O}_2 \) for indicated times (min). RNAs were divided equally and loaded directly (T: total) or methylated and isolated with anti-2,2,7-trimethylguanosine (C: capped). 6% Urea-PAGE.

Replicating the result in Figure 43, both control and flavopiridol-treated cells responded to 0.3 mM \( \text{H}_2\text{O}_2 \) with a wave of slowly growing transcripts and a restoration of 30-50 nt transcripts after 20-50 minutes (Figures 44 and 45). While the overall results are extremely similar, slight differences were observed in this repeat. In this experiment, the increased signal from short transcripts was slightly higher in flavopiridol-treated cells than the control cells after 2 minutes (Figure 45A-B); in the previous experiment, the increased magnitudes were similar with and without...
Figure 45. Profiles of total and capped nascent transcripts after H$_2$O$_2$ treatment
Whole-gel profiles for Figure 44 of total (A-B) or capped (C-D) nascent transcripts normalized using nuclear RNAs. Adherent HeLa cells were left untreated (A, C) or treated with 1 μM flavopiridol for 1 h (B, D), then with 0.3 mM H$_2$O$_2$ for the times indicated. The vertical axes represent relative phosphorimage signal intensity and all plots are scaled to have comparable no H$_2$O$_2$ (black) 30-50 nt peak heights.

flavopiridol (Figure 43B). Control cells also had more <60 nt transcripts after 5 minutes of peroxide (Figure 45A) than in the previous experiment. Finally, flavopiridol-treated cells had fewer 30-50 nt transcripts after 50 minutes of peroxide in this experiment (Figure 45B) than seen previously. Despite these subtle differences, similarities remain. The short transcripts which returned after 50 minutes of H$_2$O$_2$ treatment were slightly shorter than those before peroxide treatment in both experiments and with or without
flavopiridol. Additionally, while control cells had more short transcripts after 5 minutes in this experiment than the previous one, these transcripts still decreased after 10 minutes of H\textsubscript{2}O\textsubscript{2} in a flavopiridol-dependent manner (Figure 45A), again implying that P-TEFb acts on these polymerases after H\textsubscript{2}O\textsubscript{2}. None of the previous conclusions were discredited by this replicate.

Somewhat surprisingly, the adjusted profiles of capped nascent transcripts (Figure 45C-D) strongly correlated with the profiles of total nascent transcripts (Figure 45A-B). After 2 minutes of peroxide, the increase in capped and total transcripts was proportional regardless of flavopiridol treatment. Functional capping is not required to obtain this result if termination is inhibited, as complexes that initiated and whose transcripts were capped before the peroxide insult could accumulate after 2 minutes. The same proportional increase in capped and total 30-50 nt transcripts was also observed at the 20 and 50 minute time points and these capped transcripts are most logically generated by newly initiated polymerases after resolution of this peroxide response.

These results suggest, but do not prove, that capping in cells is unaffected by 0.3 mM H\textsubscript{2}O\textsubscript{2}. If peroxide inhibited capping, capping activity could be restored as peroxide was removed from the cell over time. Because the immunoprecipitation of capped transcripts was consistent across all conditions (Figure 44, ethidium bromide), it follows that any new initiation would require new capping for the capped and total nascent transcript profiles to remain proportional (Figure 45). If capping was inhibited immediately after peroxide treatment, both initiation and termination would have to be inhibited to obtain these results. While one could argue that uncapped transcripts from newly initiated polymerases could be degraded and avoid detection by this assay, I have previously shown that such transcripts would be visible at ~20 nt (Figures 24 and 37) and would not be present in bound fractions (Figure 39). Because total and capped nascent transcripts grew at equivalent rates over 5-20 minutes of peroxide, capping is also unlikely to influence the slow rate of elongation observed in the absence of P-TEFb activity.
$H_2O_2$ alters Pol II occupancy genome-wide

A previously published study showed that Pol II accumulated downstream of both sense and antisense promoters in MRC5 cells genome-wide after 30 minutes of 0.2 mM $H_2O_2$ and this response was downregulated at some, but not all genes after 2 hours (Giannakakis et al., 2015). Because of the time points used, however, this study overlooked the early and dynamic changes I observe in my nuclear walk-ons. I performed Pol II ChIP-Seq using adherent HeLa cells left untreated or treated for 3, 10, 30, and 100 minutes with 0.3 mM $H_2O_2$. I also performed this time course using cells treated with 1 $\mu$M flavopiridol for 1 hour prior to addition of hydrogen peroxide. By crosslinking directly in media without first washing and pelleting cells, these early time points could be obtained with confidence. All ten ChIPs were performed at the same time using the same antibody and number of cells. Similar recoveries were obtained after the immunoprecipitations and library preparation steps and similar numbers of sequencing reads were obtained across all samples. While ChIP-Seq cannot be truly quantitative without carefully controlling for variations in pulldown efficiencies (Orlando et al., 2014) and without assuming that consistent amounts of the protein of interest are present between conditions, ballpark comparisons can be made between normalized datasets with similar backgrounds. To enable such comparisons from my datasets, I normalized the pileup densities to the average number of hg19-mapped sequence reads between the matched samples. Similar normalizations were performed in the previous $H_2O_2$ ChIP-Seq study (Giannakakis et al., 2015).

First, I performed metagene analyses to look at global changes of Pol II distribution around annotated genes. Unlike previous analyses, I did not normalize the areas under each curve as my pileup densities were already semi-quantitative. These metagene analyses are fundamentally similar to my prior nuclear walk-on profiles, which also used pre-corrected inputs. Rather than looking at the sum of all transcription without considering orientation, however, these metagene profiles show the directional Pol II signal around promoters. In the absence of hydrogen peroxide, Pol II occupancy peaked about 60 bp downstream of the average TSS (Figure 46A). While distinctly separate sense and antisense Pol II peaks were not observed in this experiment, Pol II was enriched between -500 and +500 bp and antisense accumulations
Figure 46. Metagene analysis of Pol II occupancy in H$_2$O$_2$-treated cells
Average Pol II ChIP-Seq occupancy in adherent HeLa cells that were untreated (A) or treated with 1 μM flavopiridol for 1 h (B), and then treated with 0.3 mM H$_2$O$_2$ for the times indicated. ChIP-Seq reads were normalized to the average number of hg19-mapped sequences and summed over the average TSS. Curves were background subtracted but not normalized. The vertical axes represent relative ChIP signal and the untreated Pol II signal in (A) was replotted in (B) for scale (dashed line).
in signal were seen at about -80 bp and -300 bp. After 3 minutes of 0.3 mM H$_2$O$_2$ exposure, Pol II occupancy increased about 1.8-fold between -3 kb and +3 kb. Additionally, an expansion was observed in the area of Pol II enrichment (now -2 kb to +2 kb). After 10 minutes, Pol II signal slightly decreased overall but importantly, more polymerases remained in the antisense direction and these now extended -3 kb upstream. This suggests that P-TEFb stimulates productive elongation preferentially in the sense direction. By 30 minutes, Pol II occupancy fell to near normal levels (14% more than untreated) and were nearly undistinguishable from control signals after 100 minutes. While the distance of Pol II spreading was similar in both HeLa and MRC5 cells after H$_2$O$_2$ (Giannakakis et al., 2015), the magnitude and timing of Pol II accumulation were extremely different between these experiments. This could be due to differences between these cell types (for example, catalase protein levels could be significantly lower in MRC5 cells than HeLa cells).

I next looked at Pol II occupancy changes caused by H$_2$O$_2$ in flavopiridol-treated cells. In agreement with a previous Pol II ChIP-Seq performed with or without flavopiridol (Cheng et al., 2012), inhibition of P-TEFb caused a 2.1-fold increase in promoter-proximal paused Pol II (Figure 46B, dashed vs. solid black lines). In flavopiridol-treated cells, Pol II occupancy near promoters increased about 1.9-fold after 3 minutes of peroxide treatment, a magnitude nearly identical to that observed in control samples. Unlike control cells, however, Pol II occupancy further increased after 10 minutes (2.6-fold vs. no H$_2$O$_2$) and a sharp antisense protrusion was observed at about -290 bp. Curiously, Pol II signal was observed about 1 kb further downstream than upstream at this time point in flavopiridol-treated cells, while the opposite result was seen in control-treated cells (Figure 46A vs. 46B, yellow lines). This result hints at the existence of a P-TEFb-sensitive barrier to antisense elongation. After 30 minutes, Pol II signals near promoters remained generally high (2.3-fold vs. no H$_2$O$_2$); occupancy increased between +450 bp and +2 kb and decreased beyond -120 bp upstream. Only after 100 minutes did Pol II levels fall to those observed in the absence of peroxide.

To address whether or not polymerase accumulation after hydrogen peroxide treatment occurred at all promoters, or only at genes with existing transcriptional activity, I generated heatmaps for all ten datasets (Figure 47). Increases and spreading
Pol II ChIP-Seq occupancy ±1.5 kb around all RefSeq genes whose TSS (dashed line) were not within 1 kb of another TSS in adherent HeLa cells that were untreated (top) or treated with 1 μM flavopiridol for 1 h (bottom), and then treated with 0.3 mM H₂O₂ for the times indicated. Genes were rank-ordered by the median Pol II signal in untreated cells (top left) over each 3 kb area. All plots use the same white-to-black scale for relative ChIP signal intensity and can be directly compared.

of Pol II occupancy were seen only around transcription start sites with high median Pol II before treatment. Importantly, the distance of Pol II spreading after peroxide did not depend on the level of preexisting Pol II; this observation is particularly visible in ChIP-Seq experiments performed with flavopiridol-treated cells (Figure 47, bottom).
Overall, these results validate my earlier nuclear run-on findings and support a model where P-TEFb acts to enable productive elongation primarily on sense-oriented polymerases after exposure to H₂O₂. P-TEFb activity could also influence antisense elongation through a transcriptional barrier (possibly the -1 nucleosome), but even with P-TEFb, these polymerases move slowly and do not appear to enter productive elongation. Because elongation was slightly biased in the sense orientation after H₂O₂ treatment in flavopiridol-treated cells, it is possible that this barrier to elongation, if it exists, is directional and could play a role in assigning directionality to promoters.

*The role of P-TEFb at promoters and enhancers*

To verify the metagene and heatmap results and test this tentative model, I looked at a number of specific promoters over the human genome. The bidirectional *HSPA5* promoter showed clear sensitivity to hydrogen peroxide and flavopiridol (Figure 48). When flavopiridol was not present, a very clear spread of Pol II occupancy was seen upstream after 3, 10, and 30 minutes of H₂O₂ treatment and the leading edge of this antisense elongation moved at about 200 bp per minute. In flavopiridol-treated cells, Pol II occupancy extended only slightly downstream, and almost no antisense elongation was observed after H₂O₂ in flavopiridol-treated cells. Notably, both sense and antisense polymerases appeared equally sensitive to flavopiridol in the absence of hydrogen peroxide. This result indicates that P-TEFb activity prevents accumulation of antisense Pol II, though this activity might not result in productive elongation. Unlike the *HSPA5* promoter, *FOS* has very little bidirectional activity and an antisense wave of elongation was not observed after H₂O₂ in the absence or presence of flavopiridol (Figure 49). Notably, H2A.Z was present upstream, but not downstream of *FOS* in suspension HeLa cells (data not shown). While no apparent change was observed in the presence of P-TEFb activity, spreading of Pol II in the sense direction was observed after H₂O₂ treatment over time when P-TEFb was inhibited and this elongation occurred at a rate nearly identical to the antisense P-TEFb-dependent elongation seen upstream of *HSPA5*. This result suggests that similar “non-productive elongation” can occur in either direction. Curiously, the amount of Pol II remained high near the *FOS* promoter 100 minutes after hydrogen peroxide exposure in the flavopiridol, but not the control dataset. Pol II over *FOS* also remained higher than untreated MRC5 cells.
Figure 48. Pol II occupancy over HSPA5 in H₂O₂-treated cells ± flavopiridol
Pol II ChIP-Seq occupancy in adherent HeLa cells ± 1 μM flavopiridol for 1 h, and then 0.3 mM H₂O₂ for the times indicated. All tracks were normalized using the average number of hg19-mapped sequence reads. HSPA5 is transcribed right to left. RABEPK is transcribed left to right.

after 2 hours of 0.2 mM H₂O₂ (Giannakakis et al., 2015). FOS has been previously described to be upregulated in response to oxidative stress (Li and Spector, 1997; Maki et al., 1992; Rao et al., 1993) and it is possible that P-TEFb inhibition resulted in the accumulation of specifically recruited polymerases.

While promoters tend to have clearly established directionalities, enhancers generally lack productive elongation in either direction. I would therefore make two hypotheses: 1) if P-TEFb can function, enhancer polymerases should slowly spread in
both direction in response to H₂O₂, and 2) if P-TEFb is inhibited, enhancer polymerases should be inhibited by nucleosomes in both direction. To test this, I first looked at the enhancer region upstream of JUNB (Figure 50). A number of Pol II peaks were seen in the region between JUNB and HOOK2 and this area was enriched in H3K4me1 and H2A.Z. In control cells, hydrogen peroxide did not cause an increase in the enhancer Pol II peak heights but did induce an overall filling of this region with Pol II signal. Because of the number of these enhancer polymerases and their close proximity to each other, I was unable to determine the rate of this spreading. In flavopiridol-treated cells, Pol II
Figure 50. ChIP-Seq occupancy around JUNB enhancers after H₂O₂ ± flavopiridol
Pol II ChIP-Seq occupancy in adherent HeLa cells ± 1 μM flavopiridol for 1 h, and then 0.3 mM H₂O₂ for the times indicated. H3K4me1, H3K4me3, and H2A.Z ChIP-Seq occupancy in suspension HeLa cells. Pol II tracks were normalized using the average number of hg19-mapped sequence reads. JUNB is transcribed left to right. HOOK2 is transcribed right to left.
Figure 51. ChIP-Seq occupancy around DDIT4 enhancers after H$_2$O$_2$ ± flavopiridol

Pol II ChIP-Seq occupancy in adherent HeLa cells ± 1 μM flavopiridol for 1 h, and then 0.3 mM H$_2$O$_2$ for the times indicated. H3K4me1, H3K4me3, and H2A.Z ChIP-Seq occupancy in suspension HeLa cells. Pol II tracks were normalized using the average number of hg19-mapped sequence reads. DDIT4 is transcribed left to right.
signal increased after H₂O₂ and aside from the peak near chromosome 19 position 12,889,000, which had significantly more H3K4me3 than H3K4me1, appreciable elongation did not occur at enhancers. This is in contrast with the polymerases associated with JUNB, which slowly elongated downstream, but not upstream. I next looked at the enhancer region upstream of DDIT4 (Figure 51). Modest increases were observed in Pol II enhancer peak heights after 3 minutes of hydrogen peroxide in the absence of flavopiridol but most Pol II signal was distributed throughout the enhancer region at all time points. Flavopiridol caused significant accumulation of paused Pol II at these enhancers and, after H₂O₂ treatment, these Pol II peaks increased in height but not width except at chromosome 10 position 74,021,000—this enhancer also had high H3K4me3 and low H3K4me1. These results suggest that there are at least two types of enhancers and their polymerases behave differently after peroxide treatment. All enhancers, however, are flavopiridol-sensitive, which indicates that P-TEFb normally acts at enhancers.

Discussion

Studying the effects of hydrogen peroxide has led to interesting insights regarding the events of early elongation. The rapid initial increase in Pol II accumulation near promoters and enhancers is best explained by inhibition of termination after hydrogen peroxide treatment, which implies that turnover of paused elongation complexes may be more rapid than previously thought (Henriques et al., 2013). In cells only treated with H₂O₂, polymerases downstream of promoters elongated at a high rate and this transcription is likely (but not yet proven) to be the result of a transition into productive elongation. Flavopiridol treatment, which increases Pol II signal near promoters by ChIP-Seq, slowed but failed to stop this elongation. Because polymerases bound by DSIF and NELF also elongate slowly in vitro (Renner et al., 2001; Yamaguchi et al., 2013), it is possible that termination normally masks the non-productive elongation observed in these results. Although I cannot yet conclude if capping is or is not inhibited in H₂O₂-treated cells, human capping enzyme—as well as DSIF subunit Spt5 and NELF subunits A, C/D, and E—contains methionines that are oxidized by H₂O₂ (Ghesquiere et al., 2011). These effects must be part of a survival response pathway to tolerable amounts of hydrogen peroxide, though the functional
details remain unclear. Because P-TEFb-dependent elongation was observed, these effects are not due to Pol II arrest after DNA damage. It is possible that the accumulation of elongation complexes around sites of existing transcription keeps these promoters and enhancers open during brief (under 1 hour) periods of oxidative stress. This “short-term transcriptional memory” could speed a return to normal transcription after peroxide clearance.

My findings show that elongation control machinery governs the activities of Pol II at enhancers as well as promoters (Figure 52). All early elongation complexes, regardless of their origin or genomic location, appear to be affected by flavopiridol. If initiation was unchanged after flavopiridol treatment in my experiments, I also conclude that the rates of clearance—either through elongation or termination—of sense and antisense complexes at bidirectional promoter must be similar for the observed proportional increase in Pol II to occur genome-wide. At some enhancers, Pol II might enter productive elongation and these enhancers are marked by H3K4me3 and behave like directional promoters after peroxide treatment. At other enhancers, P-TEFb might induce Pol II turnover through a previously overlooked termination pathway. Curiously, P-TEFb appears to have a role in removing a barrier to elongation upstream of some genes and this is probably an H2A.Z-enriched nucleosome. Because H2O2 treatment likely inhibits termination, divergent and enhancer polymerases non-productively elongated through this barrier when P-TEFb could function—these slow transcription complexes likely lacked PAF1C or other necessary elongation factors. In the presence of both H2O2 and flavopiridol, termination and elongation through this nucleosome barrier were both blocked and this resulted in accumulation. Although P-TEFb seems to act at antisense promoters and enhancers, I suspect that additional productive elongation-associated events are required to generate the H3K4me3 mark.

Although I observed broad, global effects after hydrogen peroxide treatment, H2O2 has also been reported to induce specific gene expression (Giannakakis et al., 2015). Induction of DNA repair and stress response genes has been described as early as 15 minutes after H2O2 treatment and this induction was dependent on transcription-coupled nucleotide excision repair protein CSB (Kyng et al., 2003). Changes in expression are not detectable minutes after treatment by RNA-Seq, however, due to the
Figure 52. Resolution of bidirectional pausing with P-TEFb ± H$_2$O$_2$

The diagram illustrates the early elongation events at bidirectional promoters and enhancers. The first panel (white) depicts promoter-proximal pausing of Pol II. The second panel (teal) depicts potential functional roles for P-TEFb. The third panel (yellow) depicts changes to the results of P-TEFb activity in the presence of H$_2$O$_2$. The final panel (pink) shows the effects of H$_2$O$_2$ when P-TEFb cannot function.
abundance of existing mRNAs. A nascent transcript sequencing approach such as GRO-Seq (Danko et al., 2013; Jonkers et al., 2014) or NET-Seq (Mayer et al., 2015; Nojima et al., 2015) could instead be used to accurately quantify newly-initiated messages. This would also provide evidence regarding a potential, but unlikely \( \text{H}_2\text{O}_2 \)-induced change to the rate of initiation. Because large changes in Ser5 CTD phosphorylation were previously seen after \( \text{H}_2\text{O}_2 \) in vitro (Heine et al., 2008), EC-EMSAs could be performed on this and other CTD residues to track their changes over time (Nilson et al., 2015).

Additional ChIP-Seq experiments against DSIF and NELF would clarify whether or not non-productively elongating complexes are bound by either protein and could provide supporting evidence for productive elongation at specific genes.

Another possible purpose for this transcriptional response to hydrogen peroxide is to influence chromatin around existing sites of Pol II initiation and pausing. If changes in nucleosome stability are induced by \( \text{H}_2\text{O}_2 \) treatment, MNase-Seq and other digestion sensitivity assays can be used to identify their extent and timing. While nucleosomes take 10-15 minutes to form on DNA, prenucleosomes are established in as little as fifteen seconds and protect a region about 70-80 bp, a length distinguishable by sequencing from canonical nucleosomes (140-150 bp) (Fei et al., 2015; Khuong et al., 2016). MNase-Seq performed in cells with or without flavopiridol could also be used to verify if P-TEFb functions to loosen or remove the -1 nucleosome. Notably, stability of the -1 nucleosome has already been shown by this technique to be regulated in yeast (Kubik et al., 2015). H2A.Z ChIP-Seq with or without flavopiridol would also be informative, as resulting changes in H2A substitution dynamics should be reflected only near regions of Pol II occupancy. Finally, H3K4me3 and H3K27ac ChIP-Seq could be performed to determine if these marks are regulated during oxidative stress.

ADP-ribosylation, a post-translational modification triggered by DNA damage (Bai, 2015; Hottiger, 2015), has been recently reported to occur on chromatin in response to hydrogen peroxide (Bartolomei et al., 2016). Although PARP1, the most abundant ADP-ribosyltransferase (D'Amours et al., 1999), is thought to be enriched near TSSs (Krishnakumar et al., 2008), this peroxide-induced PARylation was found predominantly in regions of heterochromatin and was accompanied by increased MNase sensitivity at loci tested by PCR (Bartolomei et al., 2016). In addition to
nucleosomes, a large number of DNA repair proteins are also PARylated in response to H₂O₂ and knockdown of PARP1, but not PARP2, blocked this response (Jungmichel et al., 2013). PARP1 inhibition or depletion increases sensitivity of human lens cells to very low levels of H₂O₂ (30 μM) and led to more double-stranded DNA breaks, but less cell death after prolonged exposure (Smith et al., 2016). PARylation could also lead to changes in cap methylation, as both human cap methyltransferase and Fam103a1, a recently discovered HCM cofactor (Gonatopoulos-Pournatzis et al., 2011), are ADP-ribosylated (Jungmichel et al., 2013). While preliminary nuclear walk-ons indicate that PARP inhibitors reduce transcriptionally engaged Pol II in cells, this effect was not H₂O₂-specific (data not shown). It is possible that widespread PARylation occurs independently from the transcriptional response to hydrogen peroxide. Because histone PARylation lead to increased DNA accessibility, its effects should be detected by the MNase-Seq experiments proposed above without the need to perform technically complex ADP-ribosylated chromatin affinity purification (Bartolomei et al., 2016). Mechanistic studies of the effects of PARP1 on transcription in vitro would also be feasible, as HeLa nuclear extracts are unlikely to contain existing ADP-ribosylation due to its short half-life (Bartolomei et al., 2016).
CHAPTER 5: SUMMARY AND FUTURE DIRECTIONS

Overall, my research demonstrates the importance of elongation control machinery in regulating transcription from promoters and enhancers generally and in response to cellular stress. I show that the coordinated activities of initiation factors influence loading of both pausing and co-transcriptional mRNA capping machineries, which are then required for Pol II to transition into productive elongation and generate processed mRNA. After careful optimization, I performed ChIP-Seq studies in primary T cells naturally infected with HIV and discovered that STAT5A specifically binds to a sequence element within the HIV promoter after HODHBt treatment, which induced latent HIV without causing T cell activation. In contrast, my analyses of Myc occupancy in vitro and in cells demonstrates that the location of transcription machinery plays a major role in recruiting the Myc-Max heterodimer to genomic sites. A number of new questions have emerged from this work. I do not yet know what factors are associated with the non-productively elongating complexes I observe upstream of genes, at enhancers, and after hydrogen peroxide treatment; DSIF, NELF, and PAF1C are of particular interest. While antisense transcription was observed through \textit{HEXIM1}, the regulatory significance of this potential RNA interference remains to be established. I have yet to characterize the contributions of promoter proximal pausing and transcriptional interference in maintenance of HIV latency. It also remains unclear how cells benefit from the observed transcriptional response to \( \text{H}_2\text{O}_2 \). The generation of these conclusions required a unique combination of in vitro and genome-wide techniques and I believe answering these new questions will also require this marriage of biochemistry and molecular biology.

I proposed a model where H3K4me1 deposited near sites of Pol II transcription would be upgraded to H3K4me3 as a result of Set1A/B association with PAF1 complex during productive elongation. My data also suggests that P-TEFb acts at antisense promoters and enhancers and might remove an H2A.Z-enriched nucleosome barrier; however, this P-TEFb activity does not necessarily result in productive elongation. Despite the conflicting conclusions present within two recent PAF1C papers, their data supports my hypotheses. These studies contain multiple ChIP-Seq datasets that demonstrate PAF1C occupancy is predominantly downstream of promoters in
association with productive elongation complexes; PAF1C was rarely seen upstream or at enhancers (Chen et al., 2015; Yu et al., 2015). After flavopiridol treatment, Pol II paused genome-wide even if PAF1 was knocked down (Chen et al., 2015) and in the absence of productive elongation, PAF1C occupancy decreased downstream (Yu et al., 2015). Because PAF1 knockdown led to “creeping” of Pol II, the Shilatifard group concluded that PAF1 must be required for pausing; this claim is in conflict with their own Pol II flavopiridol data that showed pause release to require P-TEFb activity, not PAF1C (Chen et al., 2015). The Roeder group demonstrated that PAF1C is probably recruited by P-TEFb, as PAF1C cannot directly remove P-TEFb from the 7SK snRNP, and claimed that PAF1C facilitates release of paused Pol II into productive elongation even though changes in PAF1C occupancy after Cdk9 knockdown did not change pausing (Yu et al., 2015). I instead propose that P-TEFb, independently of PAF1C, can release paused Pol II but without PAF1C, these complexes non-productively elongate and eventually terminate. Given that PAF1C was not observed over enhancer regions in their datasets (Chen et al., 2015; Yu et al., 2015), this would explain why P-TEFb activity at enhancers does not lead to productive elongation as evidenced by a general lack of H3K4me3. The non-productive elongation described after PAF1 knockdown strongly resembles the elongation I observe in H₂O₂-treated cells at regions where PAF1 is not detected. I believe the data in these two papers support a role for PAF1C in maintenance of productive elongation and in its absence, some polymerases are released into non-productive elongation but most remain paused.

I demonstrated that Myc occupancy over the human genome is predominantly near transcription machinery and not with E-box sequence elements (Guo et al., 2014a), a finding that remains controversial (Kress et al., 2015). Notably, a very recent study demonstrated that when lysine residues within Myc are mutated, which would prevent ubiquitination and subsequent degradation, it still binds upstream of TSSs but fails to stimulate RNA synthesis or Pol II elongation (Jaenicke et al., 2016). This surprising result was demonstrated by ChIP-Seq, RNA-Seq, Western blots, BrdU incorporation assays, and both ChIP- and RT-qPCR and was validated by proteasome inhibition independent of mutation (Jaenicke et al., 2016). Because Myc interacted indirectly with CTR9 and directly with CDC73, both of which are PAF1C subunits, and this interaction
was abolished by proteasome activity, the authors concluded that Myc sequesters PAF1C just upstream of its target genes and turnover of this complex is required to induce productive elongation. This model is supported by upstream PAF1C accumulation at highly expressed genes (Chen et al., 2015; Yu et al., 2015) and explains how Myc would function when associated with elongation complexes near expressed genes (Guo et al., 2014a). However, Myc also associates with elongation complexes at enhancers (Lin et al., 2012a) where PAF1C is poorly detected (Chen et al., 2015; Yu et al., 2015). Myc activity at these enhancers is likely linked to its association with elongation complexes as disruption of these complexes by THZ1 (Nilson et al., 2015) impairs Myc-driven transcriptional activation through these enhancers (Chipumuro et al., 2014; Christensen et al., 2014). Transcriptional activation by Myc has been described to be stronger at promoters than enhancers (Nie et al., 2012), however, and it is possible that Myc acts at enhancers independently of PAF1C.

My research regarding the influences of Cdk7 (Nilson et al., 2015) and hydrogen peroxide on co-transcriptional capping have made it apparent that RNA processing is highly regulated. Capping and subsequent cap methylation are known requirements for co-transcriptional splicing (Lenasi et al., 2011) as well as polyadenylation and termination (Adamson et al., 2005). Current sequencing techniques rely on biotinylation and selection of existing caps (CAGE for example) (Takahashi et al., 2012) or Terminator exonuclease digestion of uncapped total RNA (Start-Seq) (Nechaev et al., 2010; Scruggs et al., 2015) or nascent transcripts (GRO-Cap) (Core et al., 2014; Kruesi et al., 2013), and these approaches cannot distinguish between cap types. Previous attempts to separate guanylylated transcripts which were or were not methylated relied on nuclease-contaminated trimethylguanosine antibody preparations and required addition of non-specific RNA to competitively prevent degradation (Cowling and Cole, 2007; Jiao et al., 2013; Jiao et al., 2010; Nilson et al., 2015). Aside from my MonoQ purification of this antibody, which lacks nuclease activity (Nilson et al., 2015), one other antibody presently exists that can separate G- and m7G-capped transcripts (Cole and Cowling, 2009; Fernandez-Sanchez et al., 2009). The only other method currently capable of distinguishing between caps is thin layer chromatography, which
cannot be used as an isolation method prior to sequencing (Gonatopoulou-Pournatzis et al., 2011; Jiao et al., 2010).

I believe that the regulation of capping and cap methylation can be studied global by combining my cap status determination method (Nilson et al., 2015) with nuclear walk-ons performed with biotin-11-NTPs. I should be able to quantitatively separate nascent transcripts whose caps possess or lack methylation and sequence these populations. This nascent cap status approach (which I currently call NasCap) could easily be used to address questions regarding Myc and its role in regulating cap methylation (Cowling and Cole, 2010; Dunn and Cowling, 2015) by performing such experiments in Myc-inducible U2OS cells with or without doxycycline (Walz et al., 2014). These results could be further validated by using Myc (Rahl et al., 2010) or proteasome inhibitors (Jaenicke et al., 2016) to disrupt Myc activation. Open questions also remain regarding the processing of enhancer RNAs (Li et al., 2016), whose capping has been detected to variable degrees by CAGE and GRO-Cap (Andersson et al., 2014; Core et al., 2014). Follow up experiments with THZ1, which disrupts capping (Nilson et al., 2015) and super-enhancer-induced transcription (Christensen et al., 2014; Kwiatkowski et al., 2014) could be performed in parallel to studies involving Myc, which works through these super-enhancers (Chipumuro et al., 2014; Rahl and Young, 2014). Continuing my collaboration with Alberto Bosque and Vicente Planelles, I could also use NasCap to study capping of HIV transcripts during latency and reactivation (Ferron et al., 2012; Wilusz, 2013). Previous studies have shown that viral capping is regulated by Tat (Chiu et al., 2002; Zhou et al., 2003) and subsequent cap methylation is required for expression of Rev-dependent genes (Yedavalli and Jeang, 2010).

Over the past decade, falling sequencing prices have led to a dramatic increase in the performance of genome-wide experiments and use of such techniques is all but required to publish in the field of transcription. While in vitro methods were instrumental in resolving nuanced questions regarding the regulation of capping during early elongation, I know that sequencing will eventually be necessary to show how these mechanisms function in cells and if they are differentially regulated between different genes, between promoters and enhancers, or between different conditions. Clear datasets can lead to big, high impact conclusions that permanently alter the
course of future research—for example, recent results regarding bidirectional transcription (Core et al., 2014; Scruggs et al., 2015). Resistance to sequencing techniques is not without cause. Assumptions and oversights exist in many protocols that can alter downstream results, as I found when optimizing formaldehyde crosslinking conditions for ChIP-Seq. As another example, the 5′ cap structure must be removed from nascent transcripts to enable adapter ligation during sequencing library preparation—omission of such a tobacco acid pyrophosphatase (or RNA 5′ pyrophosphohydrolase) step critically prevented detection of capped transcripts associated with promoter-proximal paused polymerases by NET-Seq in recent studies (Nojima et al., 2016; Nojima et al., 2015). In the absence of an obvious result, genomic data can be over-interpreted to support nearly any conclusion (Chen et al., 2015; Yu et al., 2015). Sometimes, honest mistakes can happen when analyzing otherwise good data (Siebert and Soding, 2014; Venters and Pugh, 2013, 2014). Moving forward, I believe that scientists must find a balance between these small- and big-picture approaches. With careful consideration for every step, I believe that results from sequencing techniques can be as rigorous and quantitative as in vitro experiments, especially when such findings are followed up and carefully validated. Such a combination of biochemistry and molecular biology can improve the quality of findings in any research endeavor.
REFERENCES


HISTONE AND tRNA MODIFICATION COMPLEXES ON ADAPTATION TO STRESS. J Biol Chem 291, 790-799.


He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N.J., Alber, T., and Zhou, Q. (2010). HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a


Hsu, P.L., Yang, F., Smith-Kinnaman, W., Yang, W., Song, J.E., Mosley, A.L., and Varani, G. (2014). Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD. J Mol Biol 426, 2970-2981.


McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev 11, 3306-3318.


152


mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell 146, 92-104.


